

# Journal of Educational Psychology

RAYMOND G. KUHLEN, EDITOR  
*Syracuse University*

## Advisory Editors

JOHN B. CARROLL <i>Harvard University</i>	CARSON MC GUIRE <i>University of Texas</i>
LEE J. CRONBACH <i>University of Illinois</i>	PAUL H. MUSSEN <i>University of California, Berkeley</i>
FRANCIS J. DiVESTA <i>Syracuse University</i>	T. ERNEST NEWLAND <i>University of Illinois</i>
CHESTER W. HARRIS <i>University of Wisconsin</i>	H. H. REMMERS <i>Purdue University</i>
ERNEST R. HILGARD <i>Stanford University</i>	JULIAN C. STANLEY <i>University of Wisconsin</i>
NICHOLAS HOBBS <i>George Peabody College for Teachers</i>	J. B. STROUD <i>State University of Iowa</i>
IRVING LORGE <i>Teachers College, Columbia University</i>	DONALD E. SUPER <i>Teachers College, Columbia University</i>
DOROTHEA McCARTHY <i>Fordham University</i>	GEORGE G. THOMPSON <i>Ohio State University</i>

---

## VOLUME 52, 1961

---

ARTHUR C. HOFFMAN, Production Manager  
HELEN ORR, Promotion Manager



PUBLISHED BIMONTHLY BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 SIXTEENTH STREET N.W.  
WASHINGTON 6, D.C.

Copyright 1961, by the American Psychological Association, Inc.

## CONTENTS OF VOLUME 52

AIKEN, LEWIS R., JR., AND DREGER, RALPH MASON. The Effect of Attitudes on Performance in Mathematics.....	19
AMIDON, EDMUND, AND FLANDERS, NED A. The Effects of Direct and Indirect Teacher Influence on Dependent-Prone Students Learning Geometry.....	286
ASTIN, ALEXANDER W. A Re-examination of College Productivity.....	173
ASTIN, ALEXANDER W., AND HOLLAND, JOHN L. The Environmental Assessment Technique: A Way to Measure College Environments.....	308
AUSUBEL, DAVID P., AND FITZGERALD, DONALD. The Role of Discriminability in Meaningful Verbal Learning and Retention.....	266
BENDIG, A. W. See MEYER, WILLIAM J.	
BLUM, STUART H. The Desire for Security: An Element in the Vocational Choice of College Men.....	317
BONKOWSKI, ROBERT J. See PIMSLEUR, PAUL	
BOSTRUM, ROBERT N., VLANDIS, JOHN W., AND ROSENBAUM, MILTON E. Grades as Reinforcing Contingencies and Attitude Change.....	112
BRADWAY, KATHERINE P., AND ROBINSON, NANCY M. Significant IQ Changes in Twenty-Five Years: A Follow-Up.....	74
CAMPBELL, ROBERT E., AND HORROCKS, JOHN E. A Note on Relationships between Student and Parent Minnesota Teacher Attitude Inventory Scores.....	199
CARTER, LAMORE J. See STROUD, JAMES B.	
COULSON, JOHN E. See SILBERMAN, HARRY F.	
CROOK, ROBERT. See TAYLOR, MARVIN.	
DAMRIN, DORA E. See RUNKEL, PHILIP J.	
DREGER, RALPH MASON. See AIKEN, LEWIS R., JR.	
DROPKIN, STANLEY. See TAYLOR, MARVIN.	
DYER, DOROTHY W., AND HARCUM, E. RAE. Visual Perception of Binary Patterns by Preschool Children and by School Children.....	161
ESTAVAN, DONALD. See SILBERMAN, HARRY F.	
FITZGERALD, DONALD. See AUSUBEL, DAVID P.	
FLANDERS, NED A. See AMIDON, EDMUND.	
GLASGOW, GEORGE M. The Effects of Manner of Speech on Appreciation of Spoken Literature.....	322
GOTTLIEB, DAVID. American Graduate Students: Some Characteristics of Aspiring Teachers and Researchers.....	236
GUMP, PAUL V. See KOUNIN, JACOB S.	
HALLWORTH, H. J. Teachers' Personality Ratings of High School Pupils.....	297
HARCUM, E. RAE. See DYER, DOROTHY W.	
HARMON, JUDSON A. See TORRANCE, E. PAUL.	
HOLLAND, JOHN L. Creative and Academic Performance among Talented Adolescents..	136
HOLLAND, JOHN L. See ASTIN, ALEXANDER W.	
HORROCKS, JOHN E. See CAMPBELL, ROBERT E.	
HUTT, VALORIE H. See RUNQUIST, WILLARD N.	
KATES, SOLIS L., KATES, WILLIAM W., MICHAEL, JAMES AND WALSH, TERRENCE M. Categorization and Related Verbalizations in Deaf and Hearing Adolescents.....	188
KATES, WILLIAM W. See KATES, SOLIS L.	
KAZMIER, LEONARD J. Cross-Validation Groups, Extreme Groups, and the Prediction of Academic Achievement.....	195
KLAUSMEIER, HERBERT J., AND LOUGHLIN, LEO J. Behaviors during Problem Solving among Children of Low, Average, and High Intelligence.....	148
KOPPITZ, ELIZABETH M., MARDIS, VERDENA, AND STEPHENS, THOMAS. A Note on Screening School Beginners with the Bender Gestalt Test.....	80

KOUNIN, JACOB S., AND GUMP, PAUL V. The Comparative Influence of Punitive and Nonpunitive Teachers upon Childrens' Concepts of School Misconduct.....	44
KRULEE, GILBERT K. See NADLER, EUGENE B.	
LOCKARD, JOAN, AND SIDOWSKI, JOSEPH B. Learning in Fourth and Sixth Graders as a Function of Sensory Mode of Stimulus Presentation and Overt or Covert Practice..	262
LOUGHLIN, LEO J. See KLAUSMEIER, HERBERT J.	
MFEE, ANNE. The Relation of Students' Needs to Their Perceptions of a College Environment.....	25
MC GUIRE, CARSON. Sex Role and Community Variability in Test Performances.....	61
MC GUIRE, CARSON. See VELDMAN, DONALD J.	
MARDIS, VERDENA. See KOPPITZ, ELIZABETH M.	
MEDINNUS, GENE R. The Relation between Several Parent Measures and the Child's Early Adjustment to School.....	153
MELARAGNO, RALPH J. See SILBERMAN, HARRY F.	
MEYER, WILLIAM J., AND BENDIG, A. W. A Longitudinal Study of the Primary Mental Abilities Test.....	50
MICHAEL, JAMES. See KATES, SOLIS L.	
MINER, JOHN B. On the Use of a Short Vocabulary Test to Measure General Intelligence.....	157
MITCHELL, JAMES V., JR. An Analysis of the Factorial Dimension of the Achievement Motivation Construct.....	179
MUEHL, SIEGMAR. The Effects of Visual Discrimination Pretraining with Word and Letter Stimuli on Learning to Read a Word List in Kindergarten Children.....	215
NADLER, EUGENE B., AND KRULEE, GILBERT K. Personality Factors among Science and Technology Freshmen.....	223
OTTO, WAYNE. The Acquisition and Retention of Paired Associates by Good, Average, and Poor Readers.....	241
✓PARNES, SIDNEY J. Effects of Extended Effort in Creative Problem Solving.....	117
PECK, ROBERT F. See VELDMAN, DONALD J.	
PIAGET, JEAN. The Genetic Approach to the Psychology of Thought.....	275
PIMSLEUR, PAUL, AND BONKOWSKI, ROBERT J. Transfer of Verbal Material across Sense Modalities.....	104
ROBESON, HAYNE W. Manifest Anxiety and Achievement Test Performance.....	132
ROBINSON, NANCY M. See BRADWAY, KATHERINE P.	
ROSENBAUM, MILTON E. See BOSTROM, ROBERT N.	
ROSENFIELD, HOWARD, AND ZANDER, ALVIN. The Influence of Teachers on Aspirations of Students.....	1
RUNKEL, PHILIP J., AND DAMRIN, DORA E. Effects of Training and Anxiety upon Teachers' Preferences for Information about Students.....	254
RUNQUIST, WILLARD N., AND HUTT, VALORIE H. Verbal Concept Learning in High School Students with Pictorial and Verbal Representation of Stimuli.....	108
RYANS, DAVID G. Inventory Estimated Teacher Characteristics as Covariants of Observer Assessed Pupil Behavior.....	91
RYANS, DAVID G. Some Relationships between Pupil Behavior and Certain Teacher Characteristics.....	82
SARASON, IRWIN G. Test Anxiety and the Intellectual Performance of College Students.	201
SECHREST, LEE. See SHORE, EUGENE.	
✓SHAW, MERVILLE C. Need Achievement Scales as Predictors of Academic Success....	282
SHAY, CARLETON B. Relationship of Intelligence to Step Size on a Teaching Machine Program.....	98
SHORE, EUGENE, AND SECHREST, LEE. Concept Attainment as a Function of Number of Positive Instances Presented.....	303

## CONTENTS OF VOLUME 52

v

SIDOWSKI, JOSEPH B. See LOCKARD, JOAN.	
SILBERMAN, HARRY F., MELARAGNO, RALPH J., COULSON, JOHN E., AND ESTAVAN, DONALD. Fixed Sequence versus Branching Autoinstructional Methods.....	166
SJAH, ANWAR. See THOMAS, R. MURRAY.	
STEPHENS, THOMAS. See KOPPITZ, ELIZABETH M.	
STROUD, JAMES B., AND CARTER, LAMORE J. Inhibition Phenomena in Fast and Slow Learners.....	30
STROUD, JAMES B. See WOLF, WILLAVENE.	
SUPER, DONALD E. Consistency and Wisdom of Vocational Preference as Indices of Vocational Maturity in the Ninth Grade.....	35
TAYLOR, MARVIN, CROOK, ROBERT, AND DROPKIN, STANLEY. Assessing Emerging Leadership Behavior in Small Discussion Groups.....	12
THOMAS, R. MURRAY, AND SJAHL, ANWAR. The Draw-A-Man Test in Indonesia.....	232
✓TORRANCE, E. PAUL, AND HARMON, JUDSON A. Effects of Memory, Evaluative, and Creative Reading Sets on Test Performance.....	207
VELDMAN, DONALD J., PECK, ROBERT F., AND MCGUIRE, CARSON. Measuring the Value Systems of Education Professors.....	330
VLANDIS, JOHN W. See BOSTROM, ROBERT N.	
WALSH, TERRENCE M. See KATES, SOLIS L.	
WASHBURN, WILBUR C. Patterns of Self-Conceptualization in High School and College Students.....	123
WEBSTER, STATEN W. The Influence of Interracial Contact on Social Acceptance in a Newly Integrated School.....	292
WOLF, WILLAVENE, AND STROUD, JAMES B. The Contribution of Response in Mental Measurement.....	249
ZANDER, ALVIN. See ROSENFIELD, HOWARD.	

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

February 1961

Number 1

## THE INFLUENCE OF TEACHERS ON ASPIRATIONS OF STUDENTS<sup>1</sup>

HOWARD ROSENFIELD AND ALVIN ZANDER

*University of Michigan*

Early studies of goal setting have shown that a person usually selects a level of aspiration which represents a mild challenge for him, given that he has knowledge of his past performances (Lewin, Dembo, Festinger, & Sears, 1944). In general, he attempts to reach the most rewarding goal that he feels he can reasonably attain. It is also known that aspirations are often influenced by sources external to the persons who set them (Festinger, 1942). Little evidence however is available concerning the nature of this influence process.

The interest of this study in the effects of teachers on a student's aspirations was stimulated by the above problem and by the awareness that appropriate goal setting by a student is an important practical issue. A salient dimension on which a pupil can hardly avoid setting a level of aspiration, and can hardly avoid being influenced by his teacher when doing so, concerns his grades in school. We define a student's level of aspiration as that level of achievement, indicated by a grade, which a pupil realistically expects to attain in a given course.

For the purposes of this investigation, we assume that teachers attempt to influence pupils to work up to their capacities, and that these pressures

are guided by professional norms that students should not be expected to work beyond or below their capacities. We assume that any given student is reasonably confident he knows the level of his best possible performance (his capacity) and that he believes his teacher knows his capacity equally well. We further assume that students are aware of the pressures from teachers and that students believe they should not be pressed to aspire either beyond or below their capacities.

The attempts of teachers to influence students are usually based upon assumptions as to what will motivate pupils to accept these inductions, such as rewards, punishments, provision of relevant information, and so on. French and Raven (1959) have proposed five separate bases of social power whose effectiveness depends upon the degree that they stimulate forces in the recipient of the influence attempts to act in accord with these inductions, minus the degree they generate forces in the recipient to resist these inductions. A summary of studies on the consequences of these different forms of social power is provided by French and Raven (1959) and by Cartwright (1959). The hypotheses considered in the present investigation are in large part suggested by French and Raven. The derivations of these hypotheses are described by those authors and will not be discussed here.

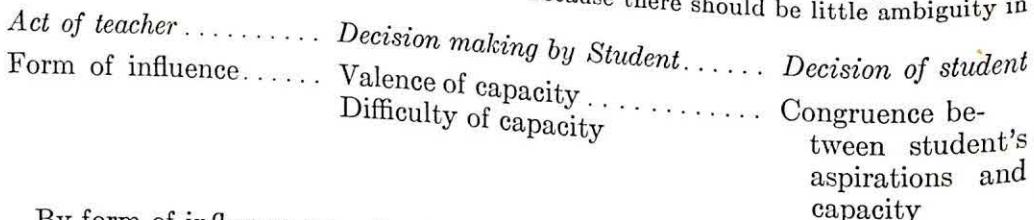
A secondary concern of this study is

<sup>1</sup> The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

in the students' perceptions of the valence and the probability of performing up to his capacity, and how these variables intervene between influence attempts by teachers and the setting of aspirations by students. This interest in the consequences of perceived valence and probability stems from assumptions in Lewin's theory of aspiration setting that the level of aspiration is a function of the positive valence of succeeding, the negative valence of failing, and the probability of succeeding or failing. Probability is measured here in terms of the perceived difficulty of achieving a capacity performance.

Finally, we consider the effects of different forms of social power upon attitudes toward teachers and the subject matter of the course. Hypotheses tested here were, for the most part, proposed by French and Raven.

The present theoretical orientation may be summarized in the following model, the terms of which are explained below.



By form of influence we refer to the bases of social power proposed by French and Raven: reward, coercion, expert, legitimate, and referent. Each of these is defined in the presentation of the results. Valence of capacity is the degree that a student perceives the attainment of his capacity grade as an attractive goal. Difficulty refers to his perception of the probability that the goal is attainable. It is assumed that the degree of difficulty perceived by a student is determined not only by his personal ability, but also by external barriers such as the competence of

teachers and the sufficiency of time and help in doing assignments. Degree of congruence between aspired grade and capacity grade is the distance between the aspired grade and the capacity grade, divided by 1.

### METHOD

Due to the exploratory nature of this research and the large number of concepts involved, a written questionnaire was selected as the basic instrument. Although it is difficult to specify with confidence the direction of causality in our correlational results, hypotheses and empirical evidence developed in other settings can be used in interpreting the most probable direction of causality. Conclusions will be stated in terms of hypotheses deemed worthy of study under more controlled conditions.

The relevant concepts were measured with questions utilizing Likert-type scales. Preliminary versions of this questionnaire were developed and revised on the basis of intensive interviews with students. The questionnaire focused on the aspirations of students in mathematics classes and on the relationship of students to teachers of mathematics. Mathematics was selected since this course is required of all respondents and because there should be little ambiguity in

students' minds over the nature of a good performance compared to courses using more subjective criteria for evaluation of their progress. The questions were repeated for English courses in order to examine the effects of a different course content and the consequences of the greater social emphasis upon achievement in mathematics which supposedly characterizes contemporary society.

Four-hundred male students, 100 from each of four junior high schools comprised the sample. Tenth graders were selected because they include the oldest group with a wide distribution of ability. The four schools were chosen to provide a wide range of socio-economic status and ability of students. A comparison of results for the predicted rela-

tionships within each of the four schools revealed that no important differences exist among the schools insofar as the present data are concerned.

Respondents in each school filled-out the questionnaire during a 1.5-hour period while teachers were absent from the testing room. The administration of the questionnaire uniformly occurred shortly after students had received their grades for the fourth of six marking periods. The recent knowledge of their grades provided the students with stable evidence of their present level of performance while permitting the possibility of future changes in the grades and in the students' aspirations for grades.

### *Validation of Assumptions*

The interpretation of results in this study rests upon assumptions stated earlier about teachers' intentions and students' perceptions of teachers' acts. The results, on the whole, support the reasonableness of these assumptions.

It was assumed that students believe their capacity is known by teachers. Evidence suggesting that this assumption is warranted for the purposes of this study is provided by the pupils' answers to the query: "Do you think your teacher is a good judge of your ability?" In the responses 67% of the students described their teachers as "quite good" or "very good" judges of their abilities. It was also assumed that students are aware of their own capacities and that they support the norm that they should not be expected to perform at levels beyond capacity. Support for these assumptions is found in responses to the question: "How reasonable is your teacher in how well he expects you to do?" Seventy-nine percent of the students replied that their teacher "expects about the right amount from me," 14% felt that the teacher required "too much," and 7% answered "too little." These percentages also provide indirect support for the assumption that teachers are guided by professional norms holding that students should not be expected to work beyond or below capacity.

A further assumption was that teachers expect pupils to work up to their capacity level. Over 80% in each of the schools perceived their teacher as expecting they could perform at the level of their capacity, while less than 1% said that they had no idea what their teacher expected of them. Most students, however, did not view these as strong demands. In response to the question: "Has

your math teacher 'pushed' you to work toward this [capacity] grade?" 55% said that they felt "little or no pressure" while only 14% said that they felt "quite a lot" of pressure.

The aspirations of students may be strongly determined by their confidence in themselves. We felt it was important, therefore, to determine what effect this personality characteristic might have upon students' attitudes toward influence attempts by teachers as well as students' aspirations and achievements. A standardized measurement of test-anxiety prepared by Mandler and Sarason (1952) was administered to all subjects. High scores on this measure indicate a high fear of failure.

While test-anxiety was not correlated with students' perceptions of teachers' influence, it was related to aspiration setting in a way consistent with findings in previous research on the level of aspiration. Atkinson has found that persons with high test-anxiety tend to avoid moderate risks (Atkinson, 1957). The present data show that students with high test-anxiety set aspirations farther from their present level of performance than do students with low test-anxiety. The goals set by the more anxious students were often unrealistically high. The high goals are taken to be unrealistic since it was noted that the farther students set their aspirations from their current grades in mathematics, the less they were likely to attain their levels of aspiration in their final grades for the course ( $r = -.63^{**}$ ).<sup>2</sup>

## RESULTS

How much were the students committing themselves when they stated their levels of aspiration? To answer this question, aspired grades were compared with the grades actually received by students at the end of the semester. The substantial relationship between aspired grades and those received at the end of the semester ( $r = .66^{**}$ ) suggests that the aspirations were realistic among a majority of students. The aspired grades usually were set from one-third to one whole letter

<sup>2</sup> Correlation coefficients reported in text and tables are marked by asterisks to indicate the probability values at the .05 (\*) and .01 or less (\*\*) levels of significance, two-tailed test.

grade higher than grades received for immediately past performances, a typical phenomenon in setting aspiration levels indicating a desire for future improvement.

Our principal concern is in the relationships of the separate forms of power exerted by teachers (as students view these matters) with the congruence between students' stated aspirations for grades and their perceived capacities.

Perceived level of capacity was measured by the query:

Not everyone can get A for a final grade in mathematics. Many students know that they must get something less because everyone has an upper limit to his ability. What final grade do you think you could get if you worked to the limit of your ability and did the best you could in mathematics for the rest of the semester?

Aspired grade was obtained by the query:

Students do not always feel like doing their best in a certain class. Sometimes they are willing to accept a grade which is not as good as they *could* get if they really tried. The final grade you will get in your math class this semester will depend partly on how hard you are going to work for the rest of this semester. Considering how hard you plan to work, what final grade do you think you should get in math this year?

#### *Effects of Separate Bases of Power*

The most direct attempts to influence students are based on the use of sanctions: by rewarding or coercing. Rewards are given or promised for behavior which is in accord with the wishes of the inducer. Coercion, based upon the ability to punish, is exerted or threatened for behavior that is not in accord with the wishes of the inducer. In a school, a teacher may administer sanctions in many ways: by the grades he gives, by comments or signs made to students, by exclusion from the group

TABLE 1  
CORRELATIONS BETWEEN FORMS OF  
POWER AND CONGRUENCE  
(N = 415)

Form of power attributed to teacher	Congruence of student's aspirations and capacity
Comparative degree of rewarding sanction	.21**
Discriminate reward, frequency	.07
Indiscriminate reward, frequency	.03
Discriminate coercion, frequency	.09
Indiscriminate coercion, frequency	-.14**
Legitimacy of grading	.21**
Expertness in grading	.09
Referent status of teacher	.13**

\*\*  $p < .01$ .

or assignment of responsibilities in class, by reports to authorities or parents, and the like. In order to encompass this variety of approving or disapproving cues, it was found necessary to cast questions about sanctioning acts by teachers in terms of the approval or disapproval that students perceived teachers have toward them. A general measure of the degree that sanctioning was perceived as rewarding or coercive was sought with the query: "On the whole, how much do you feel that your math teacher is pleased (or displeased) with you compared to the rest of the class?" It was expected that the greater the relative approval received by the student, the greater would be the congruence between his aspired grade and his capacity grade. In the first row of Table 1 it can be seen that this expectation is supported.

Although reward and coercion can be conceived as the extreme ends of a single dimension, French and Raven note that "the distinction between these two types of power is important

because the dynamics of them are different." Extreme punishment of a person, for example, may lead him to avoid or escape the whole situation in which it operates, while receipt of a valued reward may make the situation more attractive to him. Thus, measures were made of the degree that teachers reward ("Does your math teacher seem to be pleased when you do your best?") and of the degree that teachers coerce ("Does your math teacher seem displeased when you don't try very hard and your work is not as good as it could be?"). Each of these questions was answered on a frequency scale. Because the reward and coercion in these instances are being given where they are ordinarily taken to be appropriate reactions, they are designated as discriminant sanctions to distinguish them from indiscriminate sanctions, discussed in a moment. On the second and fourth lines of Table 1 are shown correlations between frequency of discriminant sanctions and congruence. The low and nonsignificant correlations suggest that the degree of discriminant reward or coercion that teachers were perceived to use do not affect the degree of congruency between aspirations and capacity.

Two further concepts relevant to sanctioning behavior were investigated. The first is indiscriminate reward: "Is your math teacher ever pleased with your work even when you don't try hard?" The second is indiscriminate coercion: "Does your math teacher ever seem to be displeased with you even when you do your best in class?" It can be seen in Table 1 that indiscriminate reward has no apparent effect upon congruence while indiscriminate coercion is inversely related to congruence.

This last finding suggests that indiscriminate coercion arouses stronger tendencies to resist the teacher's in-

TABLE 2  
CORRELATIONS BETWEEN FORMS OF SANCTION AND POWER, AND DESIRE TO CONFORM AND LEGITIMACY  
(N = 415)

Form of power attributed to teacher	Desire to conform	Legitimacy
Discriminate coercion	.13**	.06
Indiscriminate coercion	-.19**	-.24**
Difference**		
Discriminate reward	.29**	.31**
Indiscriminate reward	-.11*	-.03
Difference**		

\*  $p < .05$ .

\*\*  $p < .01$ .

ductions, than to accept them. To test such an hypothesis, students were asked: "How often do you feel like doing the things your math teacher wants you to do?" Results relevant to this hypothesis are shown in the first column in Table 2. It is plain that students perceived themselves as less ready to conform to a teacher's desires when coercion was indiscriminate than when it was discriminant.

It has been found in other research that resistance to influence becomes greater as legitimacy of influence decreases (see French and Raven). The legitimacy of social sanctioning stems from the perception, in those being influenced, that the influencer is behaving in accord with the internalized values of the ones being influenced. The degree of legitimacy attributed to teachers' behavior was measured by the query: "How fair is your mathematics teacher about most things?" The correlations between legitimacy and the separate forms of sanctioning are shown in the second column of Table 2. It can be seen that discriminant reward is reliably related to legitimacy but indiscriminate reward has no relationship with legitimacy.

Furthermore, indiscriminate coercion is reliably associated with nonlegitimacy while discriminative coercion has almost no association with legitimacy. We conclude that teachers who are considered discriminative in rewards are likely to be seen as "fair," while those who are indiscriminate in coercion are likely to be seen as "unfair."

French and Raven hypothesize that coercion arouses strong resistance in the recipient of it so that the inducer's desires are not always acted upon by the recipients and, depending upon the strength of the resistance aroused, the recipients may instead be stimulated to do the opposite of what has been asked of them. This hypothesis has been corroborated by Zipf (1958) and Sampson (1960). The present findings (and those to be seen in Table 3) suggest that resistance to coercion in the school setting may more readily generate negativism when the coercion is indiscriminate.

*Legitimate* power, we have seen, stems from the perception that an influencer is behaving in accord with the values of the person being influenced. In some instances acts by a teacher may be perceived as legitimate or non-legitimate without their being direct attempts to influence the student. An example of this type of legitimacy concerns the fairness of the teacher in evaluating the student's work. The degree of this form of legitimacy attributed to teachers was measured by the question: "If you did your best in math class would your teacher actually give you the grade that describes your ability?" It was expected that students who perceived their teacher as more legitimate would tend to set levels of aspiration closer to their perceived capacities since the risk of failure from unfair treatment by the teacher would be minimal. Evidence reported in the sixth row of Table 1 supports this

prediction. Confidence, then, that one will receive the grade he earns if he works up to capacity (and not necessarily a high grade) is associated with tendencies to aspire to attain capacity.

*Expert* power is based upon the perceived reliability of the influencer's information. The more an informer is perceived as knowing what he is talking about, the more the informer is likely to influence the recipient of the information. Since we are assuming that teachers' inductions are often placed upon the student in the direction of working up to capacity, it is evident that the teacher's attempts to influence will be more acceptable if the student perceives that the teacher knows what the student's capacity is. Thus, expertness of the teacher was measured in one way with the following question: "Do you think your math teacher is a good judge of your ability in mathematics?" It was expected that greater attribution of expertness to the teacher would be associated with greater congruence between aspiration and capacity. In the realm of teaching, however, expertness is also conceived as skill in the substantive content of the course being taught. A measure of this type of expertness is the following: "How much do you think your math teacher knows about the mathematics he is supposed to teach?" Results reported in Table 1 concern only expertness in judging the ability of students, since the results with the previous measure (expertness in math) were not different in any important respects. The non-significant correlations in the seventh row of Table 1 suggest that the expertness of the teacher does not generate greater congruence between aspirations and capacity. The failure of expertness to be related to congruence might be explained by French and Raven's statement (1959) that

expert power results in primary social influences on the person's cognitive structure and probably not on other types of systems. Of course, changes in the cognitive structure can change the direction of forces and hence of locomotion, but such a change of behavior is secondary social influence (p. 163).

Referent power exists in an influencer when others desire to be like him. Students who are highly attracted to a teacher are likely to behave in ways of which he would approve, although they may not be aware of doing so. It was found offensive to students in pretests to inquire how much they desired to associate with or be like teachers. Therefore, measurement of this form of power was by means of the question: "In general, how much do you like your mathematics teacher as a person?" The prediction that greater referent power attributed to a teacher would be associated with greater congruence between aspiration level and perceived capacity was supported, as shown in the eighth row of Table 1.

The effects of the various forms of power may be summarized by noting the positive effects of rewarding sanctions, legitimate power, and referent power on the congruence between aspirations and capacity, and the negative effect of indiscriminate coercion. It is noteworthy that the first three forms are also significantly positively related to one another, indicating that they often appear simultaneously in the teacher's behavior and often supplement one another, as proposed by French and Raven. Indiscriminate coercive power, on the other hand, is negatively related to each of the other forms. The first three forms of power, we may add, are attributed to teachers more often by students who attribute high capacity to themselves, while coercive power is attributed to teachers more often by students who assign low capacity to themselves. Nevertheless,

when perceived capacity is controlled, the relationships in Table 1 are not substantially lowered and retain their statistical significance—students with high ability not differing greatly from those with low ability.

### *Student Performance*

An important consequence of social power is the degree that the separate forms of power motivate students to *perform* at the level of their aspirations. The relationship between the closeness of the aspired grades to the *actual grade* the student received at the end of the year, and the attribution of reward power to the math teacher was  $r = .24^{**}$ . This relationship is significantly greater than its relationship to coercive power ( $r = -.08$ ). Thus, positive forms of influence appear to stimulate attainment of aspirations more than do coercive forms of influence.

### *Valence and Difficulty of Attaining Capacity*

According to the theory of aspiration setting proposed by Lewin et al. (1944) one specific aspiration level, out of a number of possible alternatives, is likely to be chosen depending upon the degree that it is attractive but not too difficult to attain. We thus expected to find students placing their levels of aspiration closer to their perceived capacity the more the capacity grades were valent for them and the less achievement of them was perceived as difficult. Valence of capacity was measured by the question: "How good do you think you would feel if you did get this grade?" Difficulty was measured with the query: "How hard would you have to work in order to receive this grade?" The expectation just stated was supported: valence of capacity grade is positively related to congruence ( $r = .12^*$ ), and difficulty of attaining capacity grade is nega-

tively related to congruence ( $r = -.27^{**}$ ). It is noteworthy that valence and difficulty are positively related to each other ( $r = .26^{**}$ ), supporting the Lewin et al. (1944) and the Atkinson (1957) findings that a difficult goal tends to be more attractive than easy ones.

We had expected to find that different forms of power would have different degrees of relationship to the valence the student attributed to the achievement of his capacity grade. A teacher who rewards a student for working at capacity, for example, might generate a greater desire in him to achieve capacity than a teacher who punishes him for not doing so. The statistical relationships between each form of influence and valence of reaching capacity were, however, consistently too low to be considered reliable. But the teachers' total amount of power in attempting to influence students appears to affect students' perceptions of the valence of the capacity grade. When all forms of power are considered together in a multiple correlation with valence, the multiple correlation between power and valence of the capacity grades is  $.27^{**}$ . The nature of the contributions made by the separate forms of power, moreover, makes it appear likely that a teacher who employs several positive bases of power simultaneously, to support his inducements on a student to work up to capacity, will have greater effect upon the valence of doing so than a teacher who employs only one positive basis of power. Why and how social power can have effects upon the valence of a goal are problems worthy of future attention.

It seems reasonable that the closer a student sets his aspired grade to his perceived capacity, the more he will be satisfied in attaining this established aspiration. This contention was supported by a correlation of  $.32^{**}$  be-

tween congruence and valence of success, the latter measured by the query: "How good would you feel if you were given the grade you intend to get?"

#### *Desire to Conform*

A direct determination of the readiness of students to be influenced by teachers was sought by the use of two related concepts: perceived desire to conform and perceived desire negatively to conform. The former was measured with the question: "How often do you feel like doing the things your math teacher wants you to do?" The latter was measured by the query: "How often do you feel like doing the opposite of what your math teacher wants you to do?" All forms of power together are strongly related to desire to conform (multiple  $R = .56^{**}$ ). Desire to conform, in turn, is related to congruence ( $r = .20^{**}$ ).

The relationships between the separate forms of power and the desire to conform, or to do the opposite, are shown in Table 3. Consistent with Table 1, indiscriminate coercion was related to nonconforming desires, indiscriminate coercion and reward were negatively related to conforming desires, and all other forms of power (including expertness) were positively associated with conforming desires.

#### *Attitudes of Students toward Teacher and Course*

A final interest of this investigation was in the relations between types of power and attitudes toward relevant aspects of the social setting. Two variables were considered here. One query asked about changes in attitudes toward the teacher ("Has your opinion of your mathematics teacher changed from what it was at first?"). The other inquired about changes in attitude toward the content of the course ("Do you feel any different about mathe-

TABLE 3  
CORRELATIONS BETWEEN FORMS OF  
POWER AND MOTIVATION  
TO CONFORM  
(N = 415)

Form of power attributed to teachers	Desire to conform	Desire negatively to conform
Comparative degree of rewarding	.35**	-.24**
Discriminate reward, frequency	.29**	-.15**
Indiscriminate reward, frequency	-.11**	.09
Discriminate coercion, frequency	.13**	.00
Indiscriminate coercion, frequency	-.19**	.16**
Legitimacy of grading	.33**	-.24**
Expertness in grading	.30**	-.20**
Referent status of teacher	.53**	-.29**

\*\* p < .01.

TABLE 4  
CORRELATIONS BETWEEN FORMS OF  
POWER AND CHANGES  
IN ATTITUDES  
(N = 415)

Form of power attributed to teachers	Change in attitude toward teacher	Change in attitude toward math.
Comparative degree of rewarding	.27**	.30**
Discriminate reward, frequency	.30**	.23**
Indiscriminate reward, frequency	-.05	-.08
Discriminate coercion, frequency	.07	.01
Indiscriminate coercion, frequency	-.22**	-.21**
Legitimacy of grading	.17**	.17**
Expertness in grading	.37**	.29**
Referent status of teacher	.57**	.40**

\*\* p < .01.

matics now than you did before you took this math course?"'). Both items were scored in terms of direction of attitude change as well as intensity of change. In Table 4 correlations are presented between these two attitudes and the different forms of power attributed to teachers. The evidence indicates that each form of power affects both of these attitudes in directions similar to the effects we have seen for grade aspirations, and desires to conform, and in directions suggested by French and Raven.

#### Results from English Classes

The results we have thus far observed concerning teachers in mathematics were not completely replicated when students were queried about their English teachers. Similar correlations were found in mathematics and English when relating forms of power to desires for conformity and to attitude changes toward the teacher and the course content. Only legitimate power, however, was signifi-

cantly related to the congruence of aspired and capacity grades in the English classes. The fact that legitimate power appears to have positive effects in both mathematics and English is understandable since legitimacy was rated by the students as the most effective source of a teacher's power in response to an inquiry into the relative importance of teachers, parents, and peers as power figures.

The failure of the other bases of power among English teachers to be related to congruence, however, requires further explanation. The original reason for the inclusion of questions about English teachers was the expectation that students would be less certain of the nature of good performance in English courses than in mathematics courses. We anticipated that influence might be more effective in English classes because students would be less clear about their capacities in that subject matter and therefore less confident of appropriate aspirations for them-

selves, thus being more vulnerable to influences from teachers (cf. Festinger, 1954). Contrary to our expectations, more uncertainty about the appropriateness of their aspirations was shown in mathematics than in English, as indicated by responses to the question: "Do you think you are aiming for too high or too low a grade?" Comparisons, furthermore, of responses to questions concerning the degree that teachers reveal their reactions to good or bad performances and the adequacy of help rendered by teachers showed no differences between mathematics and English classes. We were led, therefore, to suspect that the apparent ineffectiveness of influence attempts in English classes was due not to the course material itself or to the methods of teaching, but to the motivations of students.

An indicator of student concern over performance was available in the previously mentioned measure of valence of success. In mathematics classes, all forms of influence except coercion were significantly related to the valence of successfully achieving the aspired grades. In English classes, however, only referent power showed this relationship to a significant degree. It appears, then, that the students were more eager to do well in mathematics than in English. Further weight is lent to this interpretation by responses to the question: "Which class do you think is more important for your future?" In the replies, 34 students favored English, 152 preferred mathematics, and the rest saw them as equally important.<sup>3</sup> It is interesting to note that in the literature on level of aspiration, experimental predictions are better supported the more that

<sup>3</sup> The questionnaire was administered to students during the spring of 1959, that is, during the post-Sputnik emphasis on mathematics and physical science.

subjects have ego involvement in their tasks (Lewin et al., 1944; Stotland, Thorley, Thomas, Cohen, & Zander, 1957).

#### SUMMARY AND CONCLUSIONS

A questionnaire was used to explore the effects of teachers' influences upon students' aspirations for achievement in school. Hypotheses drawn from earlier work on the differential consequences of separate types of social power were tested in a correlational analysis.

1. Tendencies to accept a teacher's influences are aroused in students who are subject to reward, legitimate, referent, or expert power; while tendencies to ignore or oppose what teachers desire are aroused in students subject to indiscriminate coercive influences.

2. With the possible exception of expert power, these tendencies affect the degree to which students set their aspired grades congruent with their perceived capacities.

3. Two forms of coercion are distinguished by students: disapproval of inadequate performance, which appears to have no effect on aspirations or future performance, and disapproval even when performance is as good as the student feels he can do, which seems to have negative effects on aspiration setting as well as future performance.

4. Two forms of reward are also discriminated by students. Tendencies to accept a teacher's influences are lowered under indiscriminate reward but increased by reward for adequate performances.

5. The positive or negative forces set up by the separate bases of power affect the favorableness or negativeness of student attitudes toward teachers and course content.

6. The separate bases of power are effective in determining aspirations to

the degree that the students are ego involved in the performances on which they are setting aspirations.

## REFERENCES

- ATKINSON, J. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, 64, 359-372.
- CARTWRIGHT, D. Power: A neglected variable in social psychology. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959.
- FESTINGER, L. Wish, expectation, and group standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, 37, 184-200.
- FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, 7, 117-140.
- FRENCH, J. R. P., JR., & RAVEN, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959.
- LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166-173.
- SAMPSON, E. An experiment on active and passive resistance to social power. Unpublished doctoral dissertation, University of Michigan, 1960.
- STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A., & ZANDER, A. The effects of group expectation and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, 54, 55-63.
- ZIPF, SHEILA. An experimental study of resistance to influence. Unpublished doctoral dissertation, University of Michigan, 1958.

(Received August 16, 1960)

## ASSESSING EMERGING LEADERSHIP BEHAVIOR IN SMALL DISCUSSION GROUPS

MARVIN TAYLOR, ROBERT CROOK, AND STANLEY DROPKIN

*Queens College, New York*

The definition of leadership in interactional terms as typified by Gibb's (1947) discussion has created a new frame of reference for the description of leadership in behavioral terms. Instead of seeking static "traits of leadership" in an individual, investigators have been seeking criteria of "leadership behavior." For the purposes of the present study leadership may be defined as the influence of an individual in interaction with other individuals within a group setting. In recent years many promising technics have been developed for studying leadership behavior. Findings reported by the Ohio State Leadership Study Group suggest an important concept, namely, that leadership acts may be thought of as "initiating structure in interaction" and "showing consideration" (Halpin, 1956a, 1956b). What appears to be of particular significance is the finding that these two dimensions may be useful in distinguishing between effective and ineffective leaders.

The Leadership Behavior Description Questionnaire (Halpin, 1956b), a paper-pencil instrument, developed by the Ohio State group is useful in obtaining descriptions of individuals already in leadership positions. Such a technic is limited in its usefulness for assessing emerging leadership. An instrument constructed for observation of emerging leadership patterns would be particularly useful in exploring the relationship between various group-situational factors and successful leadership behavior. Such assessments would call for "on the spot" observations of behavior and their categoriza-

tion according to a set of predetermined criteria.

This paper describes an attempt to observe individuals in small leaderless groups and to categorize the emerging behavior employing the two dimensions of leadership behavior mentioned above. Before it is possible to estimate the predictive validity of such an approach, it appears necessary to establish that: (a) a group of raters can be trained to agree consistently about the classification of observed behaviors, (b) the behaviors observed are stable and consistent from one situation to another, and (c) the categories employed are sufficiently independent of each other to yield pertinent information about each individual in accordance with the underlying concept of effective leadership.

### METHOD

*Sample.* The subjects were all students in the graduate division at Queens College. Each subject was observed while participating in two, 30-minute, six-man discussion situations. The discussions were held in a one-way vision room. The 1958 sample was composed of 32 subjects and the 1959 sample was composed of 37 subjects. Each year there were four trained observers, but only two of the observers were present both years.

*Raters or Observers.* The raters were faculty members at Queens College. Training sessions were held each year prior to actual observation of the subjects. Tape recordings of discussion group situations and several practice group sessions were analyzed and rated utilizing the instrument described below. The ratings compiled during the training sessions were discussed in order to clarify and categorize various types of behavior. Once the "actual" sessions started, the observers rated independently.

*The Rating Instrument.* The theoretical concept of leadership underlying the rating instrument is indicated in the introductory comments. Its major emphasis is on those overt acts of members of discussion groups which may be classified according to the categories, described by Halpin (1956a) as: "initiates structure in interaction" and "shows consideration." In some of the earlier experimental data obtained by the authors and in a later formulation by Hempill (unpublished) it was indicated that a clearer distinction should be made between attempts to initiate structure and success in this operation. In the present technic, the latter behaviors are recorded as influence acts.

In recording an individual's behavior, if the person identifies a problem, suggests some procedure for finding a solution to a problem, or opens a channel of communication for another person, he is accorded a tally in the category, "attempts to initiate structure." If there was some overt acknowledgment by another group member of the "attempt to initiate structure," the individual is accorded an influence tally. If a person identifies himself as a member of the group, supports another's point of view, or is considerate of another's feelings or attitudes, he receives a plus tally in the category, "shows consideration." If, on the other hand, he is sarcastic, caustic, or attacks another group member, a minus tally is recorded. Thus, in this area, a group member may receive both positive and negative tallies.

*Procedures.* Each of the subjects was initially assigned at random to a discussion group. In 1958, assignments to the second discussion group were based on the principle that the group members should be new to each other. In 1959, the group membership was held constant in both situations. As far as possible all conditions were standardized. Based on previous research (Wilson & Robbins, 1955), six-person discussion groups were maintained wherever possible.<sup>1</sup> Each discussion group was assigned the

<sup>1</sup> Due to scheduling difficulties it was not possible to have all groups with six persons. The majority of the groups did have this number, however, one group had four members and another seven. Research suggests that although six is an optimal number of participants, there are not statistically significant differences between groups when four or seven people are observed (Wilson & Robbins, 1955).

same discussion problem when the members were seated in the one-way vision room. Each session lasted 30 minutes.

*Scores.* Although an individual had to attempt to initiate structure (Category 1) before he could influence the structure (Category 2) these acts theoretically represent different types of behavior. The scores in Categories 1 (attempts to initiate structure) and 2 (influence in initiating structure) were obtained by summing the number of tallies recorded for each subject by each rater. The "score" for each subject in Category 3 (shows consideration) was the difference between positive and negative tallies. Rater reliability, subject consistency and stability have been estimated for each category and are reported below.

*Analysis of Data.* The analysis of variance technique was utilized to estimate the subject consistency, subject stability, and rater agreement coefficients for each leadership behavior category.<sup>2</sup> Initially coefficients of rater agreement were obtained for each problem situation as well as for both situations combined (cf. Table 1). On the basis of these coefficients of rater agreement it was decided to combine the subject's ratings within each leadership behavior category over both problem situations. These combined ratings were treated to estimate the subject consistency (reliability) and stability coefficients for each leadership behavior category. Since the problem situation conditions (e.g., changing membership) were different for the 1958 and 1959 samples the data are analyzed independently.

The coefficients of rater agreement in each situation for each leadership behavior category were estimated by the formula:

$$R_{\text{agreement}} = (S_s^2 - S_e^2)/S_s^2$$

where  $S_s^2$  equals the between subjects mean square and  $S_e^2$ , the residual error mean square. The coefficients of rater agreement for combined situations were derived by the formula:

$$R_{\text{agreement 1.2}} = (S_s^2 - S_{sr}^2)/S_s^2$$

In this formula the  $S_{sr}^2$  stands for the variance error due to subjects by raters interaction. This coefficient may be interpreted as the correlation between ratings of  $N$

<sup>2</sup> The writers wish to express thanks to Donald Medley, Division of Teacher Education, Bureau of Research, Board of Higher Education of New York City, for his help in the statistical analysis of the data.

judges and ratings of  $N$  other judges on these same performances.

The coefficients of subject consistency (reliability) for each Leadership Behavior Category was estimated by using:

$$R_{\text{cons.}} = (S_s^2 - S_{sp}^2 - S_{sr}^2 + S_e^2)/S_s^2$$

The only new term added here,  $S_{sp}^2$ , equals the variance from errors due to the subjects by problems interaction. These coefficients may be interpreted as the correlations between ratings by  $N$  judges on two problems and ratings by  $N$  judges on two other problems. In obtaining the coefficient of stability, the formula was

$$R_{\text{stab.}} = (S_s^2 - S_{sp}^2)/S_s^2$$

This may be interpreted as the correlation between ratings on two problems and ratings on two other problems by the same judges. The assumption underlying the derivations of the various formulae were based on Model II considerations as discussed by Medley, Mitzel, and Doi (1956).

## RESULTS

Coefficients of rater agreement relevant to each Leadership Behavior Category for each problem situation and for situations combined are presented in Table 1. The coefficients for "Attempts to initiate structure" ranged from .52 to .74 in 1958 for a sample of 32 subjects and from .73 to .89 in 1959 for a sample of 37. The relationship between observer ratings for the category "Success in initiating structure" ranged from .81 to .82 in 1958 and from .83 to .88 in 1959. On the "Shows consideration" category the

coefficients ranged from .73 to .92 and from .70 to .92 in 1958 and 1959, respectively. All the coefficients presented are statistically significant well beyond the .01 level of confidence. Comparing the 1958 estimates of rater agreement with those of 1959, small increments in the latter coefficients may be noted. This may be due to increased sampling, or to the fact that at least two of the judges had had an additional year of training and experience.

Subject consistency and stability within each Leadership Behavior Category over two problem situations are presented in Table 2. The consistency coefficients range from .07 to .51 in 1958 and from .58 to .92 in 1959. The coefficients pertinent to stability of performance ranged from .07 to .51 for 1958 and from .59 to .85 in 1959. In 1959 all the coefficients were statistically significant beyond the .01 level of confidence. It may be noted that in all categories the estimates of consistency and stability are considerably higher for the 1959 data than for the 1958 data. More will be said about this below.

Table 3 presents correlation coefficients pertinent to the degree of independence, or overlap, between each of the Leadership Behavior Categories. These coefficients were computed, by the Pearson product-

TABLE 1  
COEFFICIENTS OF RATER AGREEMENT FOR EACH DISCUSSION SITUATION AND  
OVER COMBINED DISCUSSION SITUATIONS

Leadership Behavior Categories	Situation 1		Situation 2		Combined	
	1958 (N = 32)		1959 (N = 37)		1958	1959
					1958	1959
Attempts to initiate structure	.71	.89	.52	.73	.74	.89
Success in initiating structure	.82	.85	.81	.83	.82	.88
Shows consideration	.73	.70	.92	.91	.87	.92

Note.—All coefficients are significant at the .01 level of confidence.

TABLE 2

COEFFICIENTS OF STABILITY AND CONSISTENCY FOR THREE LEADERSHIP BEHAVIOR CATEGORIES OVER COMBINED DISCUSSION SITUATIONS

Leadership Behavior Categories	Consistency		Stability	
	1958 <sup>a</sup> (N = 32)	1959 <sup>b</sup> (N = 37)	1958	1959
Attempts to initiate structure	.43	.58*	.37*	.59*
Success in initiating structure	.07	.74*	.07*	.77*
Shows consideration	.51*	.92*	.51*	.85*

<sup>a</sup> Group membership in each discussion situation was changed.

<sup>b</sup> Group membership in each discussion situation was the same.

\* Significant at the .01 level of confidence.

TABLE 3  
RELATIONSHIP BETWEEN LEADERSHIP BEHAVIOR CATEGORIES (COMBINED SITUATIONS) AND PERCENTAGE OF COMMON VARIANCE

Leadership Behavior Categories	r	N	Percentage of Overlap
Attempts/Success	.03	37	00
Attempts/Consideration	.00	37	00
Success/Consideration	.65	37	41

moment method, for the 1959 data only and are based on the combined ratings for each subject. Only the correlation between the Success and Consideration categories was found to be significant at beyond the .01 level of confidence ( $r = .65$ ). The coefficients between Attempts and Success ( $r = .03$ ) and Attempts and Consideration ( $r = .00$ ) suggest little or no relationship.

#### DISCUSSION

The findings of the present study may be examined in terms of the three issues raised in the introduction to this paper.

#### Rater Agreement

The data presented in Table 1 seem to support the notion that raters (observers) can be trained to observe an individual's behavior in a small group setting and agree consistently with other observers about which acts demonstrate "attempts to initiate structure," "success (or influence) in initiating structure," or "shows consideration." This finding in itself is not particularly significant, having been demonstrated any number of times in the literature (Bales, 1950; Bass, 1954). What may prove of some importance, however, is that the raters agreed about acts which are theoretically related to effective leadership as demonstrated by the Ohio State studies of leadership.

#### Independence of Leadership Categories

Although the raters agree with each other with a high degree of consistency it is important to determine the magnitude of the relationship among the three Leadership Behavior Categories. One may ask: "Are the raters actually discriminating between the various types of behavior observed?" Table 3 presents data which suggest that there is probably no overlapping between "attempts" and "consideration" and between "attempts" and "success or influence." The relationship between "success" and "consideration" indicates that there is approximately 41% overlap in variances. Thus one may conclude tentatively that each category is somewhat independent and may be measuring important attributes of effective leadership. On the basis of the Halpin (1956a) work on effective leadership it may be postulated that a portion of the relationship between "success" and "consideration" may be due to a third variable, i.e., effective leadership. It seems reasonable to

contend that the three Leadership Behavior Categories are probably independent enough of each other to be useful measures.

From another point of view, it may be argued that an individual's "attempts to initiate structure" have little to do with whether or not he shows consideration toward the other group members. On the other hand, one must not conclude that "attempts" and "success" are not related. The present scoring system minimizes the possibility of correlation between "attempts" and "influences." At present a new scoring system is being sought, but results are not yet conclusive. Comparison of the estimates of relationship between the three categories, however, seems to support the argument that "successful initiation of structure (i.e., influence on the group problem solving procedures)" is related to consideration shown by the influential individual to other group members. Although one may try to influence his peers, apparently one must also show consideration for them if he is to succeed in influencing the group problem solving procedures.

#### *Stability and Consistency*

If any instruments to proven useful for predictive purposes or for making generalizations, it is necessary to assess behaviors which are stable and consistent. Analysis of the 1959 data suggests that subject performance is recorded as reasonably stable and consistent in all three Leadership Behavior Categories by the four raters. It is interesting to note that the AIS acts are observed as less stable and consistent than acts in the other two categories. The major source of error variance influencing all of the coefficients of stability and consistency (this is true for 1958 and 1959 data) is

the subject  $\times$  problems interaction. It may be hypothesized that the major source of variation influencing the  $S \times P$  interaction term involves the constancy of the group membership and the nature of the problem presented in each situation. In the case of the success and consideration categories the coefficients of stability and consistency are increased considerably in 1959 by holding group membership constant and presenting two similar problems.

Examination of the coefficients relevant to the AIS category indicates that the magnitude of the increments is considerably less than for the Success or Consideration categories. The present design does not permit any clear explanation of these findings. It seems logical to contend, however, that a person has more cognitive control over his attempts to initiate structure than his behavior in the other two categories. The "success or influence" behavior is, in reality, a measure of the other group members' reactions to Individual A's attempts to initiate structure. Hence, Person A has little "control" over the stability of these "acts" except as he chooses not to attempt to initiate structure. The "attempts to initiate structure" acts appear to be primarily cognitively oriented, whereas, the consideration acts appear to be basically noncognitive. If this hypothesis is tenable then Person A could not control the noncognitive acts as directly as he could control the "attempts" he makes to initiate structure. On the basis of the data presented in Table 2 it seems probable that the individual who is perceived as considerate and who attempts to initiate structure during the first group session has a higher probability of success than the individual who is not seen as considerate. It seems likely that the individual who was suc-

cessful during the first group session will continue to be successful during a second situation if the group membership is unchanged and the nature of the problems is similar. On the other hand, the individual who attempts to initiate structure but is not influential (for whatever reason) must change his behavior in the next group discussion. Hence, he may either make more or fewer attempts during the next session. Thus, his "attempts" behavior would seem less stable and consistent. Such hypothesizing is, of course, extremely tentative but may indicate some of the interesting problems about which the present technic may help to provide more information.

The instrument may have considerable use for both research purposes and more pragmatic ends. For example, the classification of leadership behavior into three categories permits potential clarification of the relationship between various leadership acts and other measures. A second project might be to find the relationship between various group dimension variables and leadership behavior categories. This technic might be useful in designing selection procedures where vocational success is highly related to effective leadership behavior.

The present instrument seems to satisfy the criteria for content validity, rater reliability, and overall reliability. The next research steps are to find out whether this technic also satisfies the criteria for construct, concurrent, and predictive validity.

#### SUMMARY

Thirty-two subjects in 1958 and 37 subjects in 1959 were observed during two leaderless group discussions by four trained observers. Each subject's

leadership behavior was categorized as "attempts to initiate structure," "success in initiating structure," or "shows consideration."

The findings suggest: (a) Raters can be trained to agree consistently about the classification of leadership behaviors observed during a group discussion. (b) The behaviors in the "success" and "consideration" categories are highly stable and consistent from one discussion situation to another when the group membership is held constant and when problems of a similar nature are being discussed. When these conditions do not prevail the leadership behaviors are not particularly stable or consistent. When the group membership is held constant the attempts to initiate structure acts are not as stable or consistent as the other two types of acts. (c) The three categories seem sufficiently independent to be considered as three distinct types of observed behaviors. The correlation between success and consideration seems to suggest a mediating variable, i.e., effective leadership.

The technic is seen to have potential value for leadership—group situational—organismic variable research. Considerably more work is necessary to determine whether the instrument satisfies the criteria for predictive, concurrent, and construct validities.

#### REFERENCES

- BALES, R. F. *Interaction process analysis: A method for the study of small groups*. Cambridge, Mass.: Addison-Wesley, 1950.
- BASS, B. M. The leaderless group discussion as a leadership evaluation instrument. *Personnel Psychol.*, 1954, 7, 470-477.
- GIBB, C. A. The principles and traits of leadership. *J. abnorm. soc. Psychol.*, 1947, 42, 268.

- HALPIN, A. W. The behavior of leaders. *Educ. Leadership*, 1956, 14, 172-176. (a)
- HALPIN, A. W. *The leadership behavior of school superintendents*. Columbus: Ohio State Univer., 1956. (b)
- MEDLEY, D. M., MITZEL, H. E., & DOR, A. N. Analysis of variance models and their use in a three-way design without replication. *J. exp. Educ.*, 1956, 24, 221-229.
- WILSON, PHYLLIS C., & ROBBINS, I. An evolving technique for appraising interpersonal effectiveness. *J. teach. Educ.*, 1955, 6, 233-238.

(Received March 31, 1960)

# THE EFFECT OF ATTITUDES ON PERFORMANCE IN MATHEMATICS

LEWIS R. AIKEN, JR.

*Woman's College, University of North Carolina*

AND RALPH MASON DREGER<sup>1</sup>

*Jacksonville University*

Recent work (Dreger & Aiken, 1957; Dutton, 1954; Gough, 1954; Poffenburger & Norton, 1956; Tulock, 1957) suggests that performance in mathematics is influenced by nonintellectual as well as intellectual variables. Mathemaphobia (Gough, 1954), i.e., pronounced fears in the presence of arithmetic and mathematics, and other negative attitudes toward mathematics demand explanation. The simplest explanation is that such reactions result from experiences specific to the learning of mathematics, in particular that the manner in which significant others, viz., teachers and parents, instruct children in mathematics is the primary determinant of their attitudes toward this subject, referred to here as "math attitudes." The present investigation provides a limited test of the direct experience hypothesis of the etiology of math attitudes by studying, first, relations between selected intellectual and nonintellectual variables and math attitudes and, second, the contributions of these attitudes to the prediction of achievement in mathematics.

## HYPOTHESES

### *Relation of Math Attitudes to Achievement Measures*

Final mathematics course grades. Math attitude scores make a significant

contribution to the prediction of final grades in a mathematics course.

Mathematics achievement test changes. Math attitude scores predict gains in scores from initial to final administration of a mathematics achievement test when training has intervened.

### *Relation of Math Attitudes to Personality Measures*

Temperament. Math attitudes are unrelated to specified "general personality" variables.

Ability. Math attitudes are positively correlated with numerical ability.

### *Relations of Math Attitudes to Experiences with Mathematics*

Ratings of mathematics teachers. Math attitudes are positively correlated with subjects' ratings of former mathematics teachers.

Reported parental encouragement. (a) Math attitudes are positively correlated with subjects' reports of early parental encouragement of mathematical endeavors. (b) Math attitudes are unrelated to subjects' reports of encouragement of studying academic subjects in general.

Reported parental attitudes toward mathematics. Math attitudes are positively correlated with subjects' reports of parents' own math attitudes.

Reported traumatic experiences with mathematics. Math attitudes from favorable to unfavorable are correlated

<sup>1</sup> The authors wish to thank Lyle V. Jones of the University of North Carolina Psychometric Laboratory for his help in planning and implementing this investigation.

negatively with the number of frustrating or embarrassing situations associated with mathematics.

### METHOD

**Measures.** Paragraphs describing attitudes toward mathematics written by 310 college students were reduced to scaled items according to Likert's procedure (Edwards, 1957) to constitute the basis for the Math Attitude Scale. The final scale consisted of 10 items connoting negative attitudes and 10 connoting positive. Sample items are:<sup>2</sup>

8. Mathematics makes me feel uncomfortable, restless, irritable, and impatient.

13. I approach math with a feeling of hesitation—hesitation resulting from a fear of not being able to do math.

18. I love mathematics, and I am happier in a math class than in any other class.

5. Mathematics makes me feel secure, and at the same time it is stimulating.

Preliminary investigation using this scale attested to its reliability ( $r = .94$  for test-retest). In addition, a test of independence between the scores on the attitude scale and scores on four items designed to measure attitudes toward academic subjects in general suggested that attitudes specific to mathematics were being measured ( $\chi^2 = .80$ ,  $df = 1$ ).

Besides the main nonintellectual measure, the Math Attitude Scale, the Minnesota Counseling Inventory (Berdie & Layton, 1957) and the Intensive Personal Data Sheet were selected to assess nonintellectual variables. The MCI was chosen not only because as a group inventory it met time requirements but also because it appears to incorporate some of the better features of its well-known cousins, the MMPI and the Minnesota Personality Scale, and to assess variables expected to relate to academic performance. The IPDS, developed at the University of Southern California, was adapted for use in this

study. Intellective measures employed were three of the Differential Aptitude Tests—Verbal Reasoning, Numerical Ability, and Abstract Reasoning (Bennett, Seashore, & Wesman, 1952)—and the Cooperative Mathematics Pretest for College Students (Mathematical Association of America, 1947), high school mathematics averages, and final grades in college freshman mathematics.

**Subjects and Procedures.** On the basis of their scores on the mathematics pretest, administered during orientation week, entering freshmen at a southeastern college who elected their mathematics for the fall semester were assigned to general mathematics, intermediate algebra, or college algebra classes. Most of the analyses were carried out on the five sections of general mathematics. All the data beyond the pretest were collected during the first few meetings of the classes, except in the case of the second administration of the mathematics pretest which took place one week before the final examinations. Data were analyzed primarily by means of multiple and partial correlation and regression methods. Hypotheses were tested for males and females separately inasmuch as a pilot study had indicated the possibility of sex differences in math attitudes.

### RESULTS

#### *Math Attitudes and Achievement Measures*

**Final Course Grades.** Multiple regression analyses of the predictive value of the Math Attitude Scale were made with the 60 males and 67 females taking general mathematics. On the basis of the intercorrelations among the five intellective predictor variables and the criterion variable (final grades), regression analyses were restricted to the independent variates of high school mathematics average, DAT Verbal Reasoning, DAT Numerical Ability, and the Math Attitude Scale. The multiple correlation coefficients were .67 and .63 for males and females, respectively ( $p < .01$ ). For cross-validation, the predictor equations for students in general mathematics were

<sup>2</sup> The Math Attitude Scale has been deposited with the American Documentation Institute. Order Document No. 6545 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1  
TESTS OF SIGNIFICANCE OF PARTIAL REGRESSION COEFFICIENTS IN MULTIPLE REGRESSION OF DIFFERENTIAL APTITUDE TEST VARIABLES, MATH ATTITUDE, AND HIGH SCHOOL MATH AVERAGE ON FINAL GRADES IN GENERAL MATH

Variable	Values of <i>t</i>	
	Males ( <i>N</i> = 60)	Females ( <i>N</i> = 67)
High School Math Average	4.12**	1.23
Math Attitude Scale	1.13	3.16**
DAT Verbal Reasoning	4.07**	.91
DAT Numerical Ability	5.09**	2.01*

\* Significant beyond the .05 level.

\*\* Significant beyond the .01 level.

females. Both of these coefficients are significant beyond the .02 level. As hypothesized, Math Attitude Scale scores predicted gains in scores on the mathematics pretest.

#### *Math Attitudes and Personality Measures*

*Temperament.* As a test of the hypothesis of the unrelatedness of "general personality" (or temperament) variables to math attitudes, multiple regression analyses were carried out, including the seven part-scores of the MCI with the DAT Verbal Reasoning, Numerical Ability, and Abstract Reasoning tests, and with a criterion of Math Attitude Scale scores. The largest portion of the variance, for both sexes, was accounted for by the regression of DAT Numerical Ability on Math Attitude Scale scores (see Table 2). However, MCI Leadership, for males, was significantly correlated with math attitudes ( $r = -.21, p < .05$ ). As Table 2 indicates, there is a slight suggestion that females with good "adjustment to reality" have more positive feelings toward mathematics than those with poorer adjustment. For males, leadership qualities and positive math attitudes may be related. In either case, the significant relation could be a chance result. Considering the general lack of correlation, it may be concluded that the hypothesis of unrelatedness of temperament measures and math attitudes is confirmed, though obviously not proved.

*Ability.* Confirmation of the hypothesis that math attitudes are positively related to numerical ability is found in the significant partial correlations between DAT Numerical Ability and Math Attitude Scale scores when the effects of the other two DAT tests are partialled out. These coefficients are .23 for 87 females and .51

applied to the scores on the independent variables of students taking algebra. The obtained and predicted grades in algebra correlated .69 for the 42 males and .65 for the 20 females. Tests of significance of the partial regression coefficients in the original equations showed that, for the males, all variables except Math Attitude made significant contributions. Only Math Attitude and DAT Numerical Ability played significant roles in the predictor equations for the females (see Table 1). Thus, the hypothesis of significant contribution of math attitudes to prediction of achievement is borne out for females, but not for males.

*Achievement Test Changes.* In the part of the study relating to gains in mathematics achievement test scores, test and retest data on the mathematics pretest were obtained on 52 males and 63 females in the three mathematics courses. The partial correlation coefficients between Math Attitude Scale scores and retest scores on the mathematics pretest, partialing out the effects of initial scores on the latter, were .33 for males and .34 for



TABLE 2

ANALYSIS OF REGRESSION OF MINNESOTA  
COUNSELING INVENTORY AND DIFFEREN-  
TIAL APTITUDE TESTS ON MATH  
ATTITUDE SCALE SCORES

Source of Variation	df	MS	F
Males (N = 96)			
Regression due to 10 variables	10	907.629	4.17**
Regression due to DAT Numerical Ability	1	5482.898	25.19**
Regression due to 9 variables (omitting Numerical Ability)	9	399.265	1.83
Regression due to 8 variables (omitting MCI Leadership and DAT Numerical Ability)	8	339.930	1.56
Error	85	217.675	
Females (N = 87)			
Regression due to 10 variables	10	855.098	3.70**
Regression due to DAT Numerical Ability	1	4847.482	20.98**
Regression due to 9 variables (omitting Numerical Ability)	9	411.499	1.78
Regression due to 8 variables (omitting MCI Adjustment to Reality and DAT Numerical Ability)	8	253.612	1.10
Error	76	231.025	

\*\*  $p < .01$ .

for 96 males in the three mathematics courses. A significant difference ( $p < .01$ ) between the two coefficients suggests that for females individual differences in Verbal Reasoning and Abstract Reasoning make important contributions to the determination of attitudes toward mathematics.

#### *Math Attitudes and Experience with Mathematics*

*Ratings of Mathematics Teachers.* The first portion of Table 3 presents evi-

dence concerning the predicted positive relation between students' ratings of their former mathematics teachers and their own math attitudes. Correlations are between scores on the rating scales of the IPDS, referring to remembered characteristics of teachers, and the Math Attitude Scale scores. The plus and minus signs in parentheses after the names of the variables in Table 3 indicate high and low ends of the scales, respectively. Thus, the value of .34 for males, opposite the Patient vs. Impatient scale, evidences a significant positive relation between positive attitudes toward mathematics and reported patience in previous mathematics teachers. It is noteworthy that more of the items are significant for females than for males, though none of the differences between the coefficients for males and females is statistically significant. The number of significant coefficients probably does not arise by chance. Math attitudes are thus apparently related to remembered impressions of teachers, the female more clearly so than the male attitudes.

*Parental Encouragement.* The obtained correlations between the Math Attitude Scale and IPDS rating scale variables used in a test of the hypothesis of a positive relation between attitude and parental encouragement in mathematics and study in general are listed in the second section of Table 3. Since none of the eight correlations is significantly different from zero, there is no evidence that math attitudes are related to memory of parental encouragement of studying mathematics or academic subjects in general.

*Parental Attitudes toward Mathematics.* To obtain data on the hypothesized relation between students' math attitudes and their reports of their parents' attitudes toward mathe-

TABLE 3  
CORRELATIONS BETWEEN MATH ATTITUDE  
SCALE SCORES AND SCORES ON ITEMS  
RELATED TO STUDENT IMPRESSIONS  
OF MATHEMATICS TEACHERS AND  
EXPERIENCES WITH PARENTS

Scale	Correlation with Math Attitude Scale	
	Males (N = 96)	Females (N = 87)
Your impressions of your math teachers:		
Patient (+) vs. Impatient (-)	.34**	.26*
Strict (-) vs. Lenient (+)	-.06	-.10
Hostile (-) vs. Friendly (+)	.22*	.15
Fair (+) vs. Unfair (-)	.14	.17
Demanded high standards (+) vs. Did not care (-)	.15	.31**
Domineering (+) vs. Submissive (-)	.06	-.05
Lots of fun (+) vs. Grim (-)	.08	.27*
Brutal (-) vs. Kind (+)	.07	.21*
Clever (+) vs. Dull (-)	.10	.26*
Nervous (-) vs. Controlled (+)	-.04	.10
Knew their subject well (+) vs. Were severely lacking in knowledge of subject (-)	.13	.34**
Really knew how to teach math (+) vs. Did not know anything about how to teach math (-)	.18	.31**
As you experienced your father when you were a child:		
Stressed my school work greatly (+) vs. Paid no attention to my school work (-)	-.04	.12
Encouraged me to study math (+) vs. Discouraged me from studying math (-)	.08	.11
As you experienced your mother when you were a child:		
Stressed my school work greatly (+) vs. Paid no attention to my school work (-)	-.07	.00
Encouraged me to study math (+) vs. Discouraged me from studying math (-)	-.10	.14

\* Significant beyond the .05 level.

\*\* Significant beyond the .01 level.

matics, Math Attitude Scale scores were correlated with the scores on the following scales of the IPDS:

As you experienced your father when you were a child: Liked math vs. Disliked math

As you experienced your mother when you were a child: Liked math vs. Disliked math

The correlations between the Math Attitude Scale and the two variables above were .08 and .10, respectively, for the 96 males and .13 and .16 for the 87 females. Since these coefficients are not significantly different from zero, the hypothesized relation between students' reported perceptions of the parents' feelings toward mathematics and the students' own attitudes toward the subject is not confirmed.

*Traumatic Experiences with Mathematics.* The answers to the following questions on the IPDS were used to test the hypothesized relation between attitude and traumatic experiences with mathematics:

Can you remember any specific embarrassment or insecurity pertaining to your performance in arithmetic or math when you were a child? If so, describe. How old and in what grade were you when this took place?

A chi square test of independence between Math Attitude Scale scores, dichotomized at the median, and a Yes-No answer to the first question gave values of .62 and 1.11 for the 96 males and 87 females, respectively. Since neither value is significant, the hypothesis of independence between the two variables cannot be rejected. There were not enough responses to the remaining items to make a statistical analysis.

## DISCUSSION

Assuming that the various measures employed for this study accurately assess their respective variables, we may conclude that the direct experience hypothesis which guided the investigation overall is partly supported, but only partly.

Math attitudes are apparently related to intellective factors and achievement, but not to temperament variables, at least within the limitations of

this study. Experiences with former mathematics teachers are somewhat related to present math attitudes, although remembered parental encouragement and math attitudes and own remembered traumatic experiences are presumably unrelated.

The partly unrealistic nature of the assumption of validity of measures employed in this study is easily seen. Final grades, for example, are not entirely adequate to assess achievement in mathematics. The group temperament inventory suffers from the well-known failings of such measures. Within the limitations of the measures, however, the conclusion of the influence of direct experience upon math attitudes may be upheld.

Other research will have to be undertaken to determine if factors not investigated here are operative in determining math attitudes. At best the proportion of variance of the variables here associated with math attitude variance is small. Other variables must be uncovered whose variance further accounts for that in the Math Attitude Scale. Perhaps a less direct method of ferreting out traumatic experiences with mathematics, for one instance, would reach some of the bases of mathemaphobia and other negative attitudes. Projective measures or objective inventories other than those we have employed might reveal relations not found in this study.

#### SUMMARY

The Math Attitude Scale, DAT Numerical Ability, Verbal Reasoning, and Abstract Reasoning tests, the Cooperative Mathematics Pretest for College Students, the Minnesota Coun-

seling Inventory, and an adaptation of the Intensive Personal Data Sheet were administered to college freshmen in mathematics courses to determine in a limited way the etiology of math attitudes. Regression and correlation analyses of the intercorrelations of these measures and their relations with grades in high school and college mathematics courses supported to a modest extent the supposition that direct experiences in relation to mathematics contribute to math attitudes. Other influences in the determination of negative math attitudes are not excluded by the findings in this study.

#### REFERENCES

- BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *Differential Aptitude Tests: Manual*. New York: Psychological Corporation, 1952.
- BERDIE, R. F., & LAYTON, W. L. *Minnesota Counseling Inventory: Manual*. New York: Psychological Corporation, 1957.
- DREGER, R. M., & AIKEN, L. R. The identification of number anxiety in a college population. *J. educ. Psychol.*, 1957, 48, 344-351.
- DUTTON, W. H. Measuring attitudes toward arithmetic. *Elem. sch. J.*, 1954, 55, 24-31.
- EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.
- GOUGH, M. F. Mathemaphobia: Causes and treatments. *Clearing House*, 1954, 28, 290-294.
- MATHEMATICAL ASSOCIATION OF AMERICA, Committee on Tests. *Cooperative Mathematics Pretest for College Students: Manual*. Princeton: Cooperative Test Service, 1947.
- POFFENBURGER, T. N., & NORTON, D. A. Factors determining attitudes toward arithmetic and mathematics. *Arith. Teach.*, 1956, 3, 113-116.
- TULOCK, M. K. Emotional blocks in mathematics. *Math. Teach.*, 1957, 50, 572-576.

(Received April 27, 1960)

## THE RELATION OF STUDENTS' NEEDS TO THEIR PERCEPTIONS OF A COLLEGE ENVIRONMENT<sup>1</sup>

ANNE McFEE<sup>2</sup>

*Syracuse University*

Over the past 3 years the College Characteristics Index (CCI) (Stern & Pace, 1958) has been filled out by several thousand students in more than a hundred colleges (Pace, 1960; Pace & Stern, 1958; Thistlethwaite, 1959). The instrument is intended to give an estimate of the press of the college environment. The 30 press scales in the CCI parallel the 30 needs scales in the Stern Activities Index (AI) (Stern, 1958). In using and interpreting the instrument, it is clearly important to know whether the personality of the students who answer its items has any appreciable relationship to the way they answer them.

McConnell and Heist (1959), noting that the personality characteristics of student bodies vary widely from one college to another, and even between colleges which are highly selective in scholastic aptitude, have raised the question: "Do students make the college?" If one is to study the interaction between students and environments one must have independent estimates of each. The CCI should give an estimate of the environmental press independent of the personality needs of the students responding to it. Does it in fact do so? This study attempted to answer this question on two levels: the general relation between corresponding need and press measures, and the specific relation of each CCI

item to a relevant personality need scale.

Two other factors were studied. The first is the objectivity of the CCI item. The hypothesis here is that the more easily verifiable the behavior or knowledge the item describes, the less likely it is that people will see it in an individual way. The second factor considered is the likelihood that the student has a basis of personal experience for saying that the item is "true" or "false" of his environment. The hypothesis is similar to the one relating to the objectivity of the item, i.e., the more familiar the students are with the behavior described in the item, the more they will agree on its truth or falsity. When few students have experienced the behavior in question, there will be more disagreement.

The respondent to the CCI is asked to report whether certain specified behaviors or conditions are true of his college environment. It seems likely that this situation would tend to minimize unique, individual sets, and facilitate the expression of opinions that have been acquired by members of the college group through contact with the immediate environment. The perceptions of a particular group should show a high degree of agreement, but there should be considerable differences among groups. Preliminary study of the variance of scale scores within and between colleges on an earlier version of the CCI has shown this to be the case (Pace & Stern, 1958).

<sup>1</sup> This paper is based on the writer's MA thesis in Psychology at Syracuse University. The author wishes to thank C. R. Pace for his extensive suggestions and guidance throughout the course of the study.

<sup>2</sup> Now at Stanford University.

## METHOD

"Perception of the environment" is defined by the answers students give to the 300 items of the CCI. Briefly, the index consists of 30 scales, with 10 true-false items in each scale. The items describe commonplace activities or conditions which occur or might occur at college. The scales describe 30 types of environmental press. A press is an aspect of the environment which tends to encourage or reward a particular type of behavior. For example, if the item "Many upperclassmen play an active role in helping new students adjust to campus life" is answered "true" it is presumed to reflect a press toward nurturant behavior.

The personal characteristics of the student examined in this investigation are his scores in the AI. The AI also consists of 30 scales of 10 true-false items each, but it is designed to measure level of need in the individual, rather than intensity of press in the environment.

The items of the CCI were classified for objectivity and exposure value by three judges working together. Approximately 2 weeks later, two of the judges went over the classifications. When there was doubt about an extreme item, it was shifted toward a middle category, in an effort to make extreme categories as free as possible from ambiguities. In this second session, only about 10% of the items were shifted.

For each item of the CCI, response frequencies were tabulated for people answering in the direction of the key, and in the opposite direction. In other words, for each item, there was a group of respondents who "passed" it, and another group who "failed" it. The mean AI score, on the scale corresponding to the CCI scale in which the item appeared, was calculated separately for the "pass" and "fail" groups. For each CCI item,  $t$  tests were computed between AI scale means for the "pass" and "fail" groups. For a sample of this size, the  $t$ 's would not be significant if the differences between the means were less than .50, so  $t$  ratios were calculated only for items showing differences of .50 or above. This minimum difference was necessary for significance no matter how the "pass" and "fail" groups were proportioned (97-3, or 50-50).

*Subjects.* Responses to both the CCI and the AI were obtained from 100 students in introductory psychology classes at Syracuse University. They are not necessarily a representative sample of Syracuse students;

the study is not intended to provide a generalization about a particular university, but only about groups of college students.

## RESULTS

### *Perception Related to Personality Needs*

Pearson product-moment correlations were calculated between each pair of scales carrying the same label on the AI and CCI, e.g., between Need Achievement-Press Achievement. These correlations ranged from -.007 to .057. The median correlation was .006. Of the 30 correlations, 24 were in a positive direction, and 6 were negative.

For a sample of 100 cases, a Pearson  $r$  of .197 will be significant at the 5% level of confidence. None of these correlations is significant. Since on a chance basis alone occasional  $r$ 's of much larger magnitude than those obtained would be expected (one or two above .197) all computations were double checked. It is not likely that the lack of demonstrated relationship is due merely to low scale reliabilities. Reliabilities for the CCI and the AI were computed using Kuder-Richardson Formula 20 (Stern, 1959). These range from .34 to .81 for the CCI, with a mean of .65. For the AI they range from .40 to .88 with a mean of .69.

### *Response Uniformity Related to Objectivity and Exposure Value of Items*

To test the hypothesis that easily verifiable items will be answered more consistently by the students, the 300 CCI items were classified as (a) highly objective, easily verifiable from obvious criteria; (b) somewhat objective, verifiable from criteria requiring an observer or otherwise less obvious; and

TABLE 1

VARIABILITY OF STUDENT RESPONSES AND RELATION OF STUDENT NEEDS  
TO CCI ITEMS VARYING IN POSSIBILITY OF OBJECTIVE VERIFICATION  
AND IN EXPOSURE VALUE

Characteristic	No. of Items	% producing uniform responses	% producing divided responses	% significantly related to needs <sup>a</sup>
<b>Objectivity of Items:</b>				
Highly Objective Items	62	13	42	8
Somewhat Objective	107	—	—	—
Subjective	131	5	65	8
<i>z</i>		2.0*	2.3*	0
<b>Exposure Value of Items:</b>				
High Exposure	51	24	41	6
Medium Exposure	119	—	—	—
Low Exposure	130	3	62	19
<i>z</i>		4.2**	2.6*	2.2*

<sup>a</sup> Significant *t*'s (.05 level) between AI scale means for "pass" and "fail" groups on each CCI item.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

(c) subjective, confirmable only by asking more people the same question. Response frequencies were compared for the three groups. The results are shown in Table 1.

There is a high degree of uniformity of response to an item if nearly everyone answers it the same way, whether this is in the direction of the key or in the opposite direction. Of the 62 items classified as highly objective, 13% had response percentages of 90 and above or 9 and below, compared to 5% of the 131 items classified as highly subjective. Thus the more objective items tend to evoke a somewhat larger percentage of highly uniform responses than do the subjective items. This difference is significant at the 5% level (*z* = 2.0).

Of the highly objective items, 42% fall in the middle response range of response percentages (30 to 70 in response percentages), and 65% of the highly subjective items fall into this response range. The middle range represents the responses that are the

least uniform, since they center on the 50%, or completely divided, level. A smaller proportion of highly objective than highly subjective items falls in this range, i.e., produce disagreements among respondents. This difference is significant at the 5% level (*z* = 2.3).

To examine the hypothesis that students will agree more in reporting behavior with which they are all likely to be familiar, the items were classified under three levels of probable familiarity or "exposure value": high, medium, and low.

Of the 51 items classified as high in exposure value, 24% had response percentages of 90 or above or 9 and below, compared to only 3% of the 130 items classified as low in exposure value. The percentage of uniform responses to items of high exposure value was eight times the percentage of uniform responses to items of low exposure value. This difference is significant beyond the 1% level (*z* = 4.2).

Of the high exposure items, 41% fell in the middle response range, and 62% of the low exposure items fell in this range. A smaller proportion of high exposure than low exposure items produced disagreement among respondents. This difference is significant at the 5% level ( $z = 2.6$ ). In this respect, there is no difference between the objective-subjective and high exposure-low exposure types of classification. Both produce the same proportion of items in the middle response range.

#### *Relation of Responses to Personality Need, Objectivity, and Exposure Value of Item*

Of the 300 CCI items, 36 (12%) showed  $t$  ratios, significant at the .05 level or above, between "pass" and "fail" group means on the corresponding AI scale.

Of the items classed as highly objective, 8% showed significant  $t$  ratios. Of the highly subjective items, again 8% showed significant  $t$  ratios. The objectivity of an item does not seem to be related to the degree of influence of personality needs on student responses.

Six percent of the high exposure items showed significant  $t$  ratios between "pass" and "fail" group means on the corresponding AI scale. Of the low exposure items, 19% showed significant  $t$  ratios. When a student, lacking direct experience, must guess at the answer to a question, whether objective or subjective in content, his own needs are more likely to influence his judgment to a significant degree, as reflected by these  $t$  ratios. The difference in proportions of significant  $t$ 's between the high and low exposure items is significant at the 5% level ( $z = 2.2$ ).

#### SUMMARY

This study attempted to clarify some of the relations between student perception of the college environment and various other factors. If the College Characteristics Index is to be useful to investigators as an objective indicator of differences between colleges, it should be independent of the personality needs of the informants filling it out. This study failed to find any correlation between scale scores of individuals on the CCI and their parallel scores on the AI, a personality test using parallel scale classifications; nor was a strong relation found between personality need and the students' perception of environmental press, as reflected by individual items. The responses to 88% of the 300 CCI items were independent of the parallel personality need of the respondent.

Differences in objectivity of individual items produced a moderate difference in uniformity of response to the items, but produced no discernible differences in the influence of need on the item responses. Items about behavior or conditions which the student is unlikely to have encountered (i.e., those low in "exposure value") produced much less agreement, and were much more influenced by need, than were items about more widely shared experience.

#### REFERENCES

- McCONNELL, T. R., & HEIST, P. Do students make the college? *Coll. Univer.*, 1959, 34, 442-452.
- PACE, C. R. Five college environments. *Coll. Bd. Rev.*, 1960, 41, 24-28.
- PACE, C. R., & STERN, G. G. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, 49, 269-277.
- STERN, G. G. *Activities Index, Form 1158*. Syracuse, N. Y.: Psychological Research Center, Syracuse Univer., 1958.

- STERN, G. G. *The ecology of the American college student: Varieties of constraint in American education.* Syracuse, N. Y.: Psychological Research Center, Syracuse Univer., 1959. (Mimeo)
- STERN, G. G., & PACE, C. R. *College Characteristics Index, Form 1158.* Syracuse,

- N. Y.: Psychological Research Center,  
Syracuse Univer., 1958.
- THISTLETHWAITE, D. L. College press and  
student achievement. *J. educ. Psychol.*,  
1959, 50, 183-191.

(Received May 2, 1960)

## INHIBITION PHENOMENA IN FAST AND SLOW LEARNERS

JAMES B. STROUD

AND

*State University of Iowa*

LAMORE J. CARTER

*Grambling College*

The purpose of this study was to investigate intralist and interlist inhibition as a function of length of list and level of ability (rate of acquisition). Two groups of subjects, widely separated in rate of learning, were compared both on learning and recall of lists of two different lengths, in which all items by all subjects were learned to a common criterion of two and only two reinforcements. By definition, slow learners require more trials, more item presentations, than fast learners. They require more presentations to achieve the initial correct responses in the course of learning a list. It is also possible that the stability of correct responses, as determined by the probability of their occurring twice in succession, once they have occurred, may vary with ability level. On long lists, all subjects require more item presentations to achieve the initial correct responses than they do on short lists (Robinson & Darrow, 1924; Robinson & Heron, 1922).

Also it is possible that the increasing of length would affect the stability of responses of all subjects, but that of slow learners more than of fast learners. Possibly fast learners behave on long lists much as do slow learners on short ones. One of the more important differences between the two ability groups may lie in a difference in their ability to withstand the effects of intralist inhibition.

As stated thus far, the experiment was so planned as to: (a) investigate the relative effects of length upon trials required to learn by two widely separated ability groups, and (b) compare recall scores of the two groups of subjects on two different lengths of

lists: under the condition in which all individual items on both lists by all subjects were learned to a common operational criterion.

As an attempt to further the purpose of this investigation, certain items just learned as a warm-up exercise were interspersed both in the short and the long list. It seemed worthwhile to know whether or not the retention of the warm-up items, when presented in a larger context, would vary with ability level and length of list.<sup>1</sup> This amounts to what appears to be a procedure for investigating retroactive inhibition as a function of length of list and of ability level.

The question of differences by ability level in susceptibility to retroactive inhibition seems rather important. In two recent investigations, no relationship was found between rate of acquisition and recall (Stroud & Schoer, 1959; Underwood, 1954). If ability differences are associated with resistance to inhibition as just suggested and if this should turn out to be a general phenomenon, then slow learners should be more susceptible to retroactive inhibition effects than fast learners, and should, in terms of widely accepted theory of forgetting, retain less well what they learn. Perhaps all this is a bit tenuous, but it does suggest some outcomes that would appear to be incompatible with the experimental data reported in the two investigations just mentioned.

<sup>1</sup> This procedure appears to be an efficient method of investigating both proactive and retroactive inhibition—particularly the latter since it very largely gets rid of the troublesome problem of rehearsal.

### PROCEDURE

Preliminary learning tasks were employed for purposes of selecting two groups of learners, one fast and one slow. A list of 12 paired adjectives and one of 10 paired picture-names were used. The latter consisted of pictures (faces and shoulders) of male college students and fictitious first and last names. These were placed on film strips and projected upon a screen before classes of college sophomores. Throughout the first trial, a picture or an adjective, as a stimulus member, was exposed for 2 seconds; then both members of a pair were simultaneously exposed for 2 seconds. When all the items in the list were thus exposed, there followed a recall trial in which the stimulus members alone were exposed, 4 seconds for the adjectives and 6 seconds for the pictures. During this time the subjects wrote in specified blank spaces such of the responses as they could recall. Two exposure and recall trials were allowed on the paired adjective list and three on the paired picture-name list. Total number of correct responses on the two lists combined constituted the scores.

By this means, two groups of 32 subjects each were selected, representing the top and bottom 15% of the group sampled.

In the experiment proper, two lists of paired adjectives of average associative and familiarity values from the Haagen (1943) list were used. Three separate lists of 12 items each were prepared. First one, then another, of these lists was used as a 12-item list, each being used an equal number of times for this purpose. The two remaining lists were combined to form a 24-item list, which was learned as such. The three blocks of 12 items were systematically rotated throughout among the 12- and 24-item lists, in an attempt to control possible differences in item difficulty between the long and short lists.

The adjective pairs were inscribed on  $4 \times 6$  plastic cards, a pair per card, appropriate for use in the Card Master. In the experiment, the first member of a given pair was exposed for 2 seconds, following which both members were simultaneously exposed for 2 seconds. As learning proceeded, subjects were required to anticipate verbally the response member within the 2-second interval in which the stimulus member alone was exposed. A trial consisted of a single exposure, in this manner, of all the items in a list. The serial order of the cards was varied systematically throughout the learning trials. Each subject learned both the long and the short list, the order alternating

from subject to subject. An interval of 4 seconds was interspersed between trials.

Each subject, just prior to undertaking each of his learning tasks, learned a warm-up list consisting of three pairs, by the method just described. Each such list was learned to a criterion of five errorless trials. For purposes stated presently, each of the three warm-up pairs, when learned, was placed in the appropriate list to be learned. Operationally this increased the length of each list by three items. On the short list the three warm-up items were placed in the fifth, tenth, and fifteenth positions; on the long list, at the ninth, eighteenth, and twenty-seventh positions.

As a further condition of the experiment each item, except the warm-up items,<sup>2</sup> was withdrawn, from the list when and as it was correctly responded to twice, consecutively or not. This procedure was adopted for two reasons: it had the effect of shortening learning time; it insured the same number of operationally defined reinforcements on all items of both lists by all subjects. The latter seemed especially important for investigation of the effect of length and ability differences upon retention.

Subjects appeared, individually, at appointed times, learned the appropriate warm-up task and the appropriate main experimental task. After an interval of 48 hours, they reappeared, engaged in a recall performance on the original experimental task, under the same conditions as those under which it was learned, and proceeded to learn the second warm-up task and the second main experimental task. For this task, a recall performance was exacted 48 hours later.

### RESULTS

The means of the number of trials required to learn the short and the long lists by the two ability groups are presented in A of Table 1. There is little room for doubt that the selection procedure used produced two groups of subjects far apart in learning ability. By the procedure of deleting items from the list as soon as they had been twice responded to successfully, trials became shorter as learning progressed. By number of trials is meant the

<sup>2</sup> The Card Master does not work well when fewer than four cards are in the machine.

TABLE 1  
LEARNING AND RECALL DATA

Task Score	Subjects	
	Fast	Slow
A. Means of Trials to Learn		
Short List	9.16	36.00
Long List	36.38	114.88
B. Mean Number of Consecutive Successful Anticipations		
Short List	11.34	9.78
Long List	21.90	18.40
C. Mean Number of Unsuccessful Anticipations of Warm-up Items		
Short List	.22	2.13
Long List	.44	2.66
D. Means of Recall Scores		
Short List	5.38	5.13
Long List	13.25	10.94

number of times a list or some part of it was presented.

Under this revised-list method, the increasing of the length of list resulted in the increasing of the number of trials required to learn and did so at an accelerating rate. This is in keeping with results on length obtained by the conventional method of learning the list as a whole to a criterion (Robinson & Darrow, 1924; Robinson & Heron, 1922).

The increasing of length resulted in a greater increase in number of trials required to learn by the slow learners than by the fast learners—78.88 to 27.22. An analysis of variance procedure yielded a significant Length  $\times$  Ability interaction ( $p < .005$ ). Incidentally, the Length  $\times$  Order interaction was also significant at a like level, suggesting that the practice effects from the long list upon the short list were greater than those of the short list upon the long one.

The data were next analyzed with respect to the number of times correct responses once made occurred twice in

succession, for the two ability groups and the two lengths. In this analysis the three interspersed warm-up items were counted, with the result that the maximum number of times two consecutive correct responses could have been made was 15 and 27, for the two lists. The results are presented in B, Table 1.

On both lists, the number of consecutive successful responses made by the fast group was somewhat greater than that made by the slow group. This is consistent with Underwood's (1954) observation that the reinforcing of a slow learner's response contributes less to habit strength than that of a fast learner. However, the relative effect of list length upon the two groups was about the same.

The mean number of unsuccessful anticipations of the three previously learned items (warm-up) interspersed in the lists is presented in C, Table 1, by ability level and by length of list.

Ability differences were significant ( $p < .005$ ). Ability  $\times$  Length interaction was not significant. The differences associated with length were not significant. It seems clear that the fast learners could better withstand the interfering effects resulting from the presentation of the warm-up items within a context of similar items than could the slow learners. It may be that with longer warm-up lists, in the sense used here, significant length effects and Length  $\times$  Ability interaction would have been obtained.

Recall scores of the two groups of subjects on the two lengths were compared. D, Table 1, shows the mean words recalled—within a 2-second exposure interval, 48 hours after learning—for these comparisons.

Investigations of the relation between list length and recall have obtained higher scores in percent recalled for the longer lists, when entire lists were learned to a common criterion

(Robinson & Darrow, 1924; Robinson & Heron, 1922). This has generally been explained in terms of a relatively greater overlearning of some of the individual items in the longer lists than in the shorter ones. In this experiment, all items were learned to the same objective criterion.

Long lists require more trials to learn, but when learned are retained as well as short lists learned in fewer trials. Consistent with this is the proposition that slow learners require more trials to learn lists of all lengths than do fast learners, but when learned retain them as well as do fast learners. The over-all ability effect was not significant. However, separate analysis gave a significant difference between the two ability groups on the long list. The Length  $\times$  Ability interaction was significant ( $p = .05$ ).

Perhaps the order effects on recall are worth mentioning in this connection. In all of the foregoing analyses, order effects (order in which the lists were learned) were determined. Generally, these were large and significant. Also, order interactions as Order  $\times$  Ability, Order  $\times$  Length were significant. In recall, there were no order effects.

### DISCUSSION

Probably no one doubts that there are differences among people in recall or in other measures of memory. In most practical life situations, differences in degree of learning, in familiarity, in subsequent utilization of learning, rehearsal, and others operate. These have an effect upon recall. There appears to be no compelling reason to posit some kind of differences inherent in basic psychological process in order to account for the observed differences. However, one possible basic difference in psychological process does suggest itself.

There is some evidence in the results

of the present investigation that slow learners are more susceptible to inter-list interference than fast learners. This suggests that they may be more susceptible to the effects of retroactive inhibition, in the traditional sense. Our findings relative to ability differences in the recall of the warm-up items (C, Table 1), we think, support this. If various kinds of interfering effects do affect slow learners more adversely than fast ones, and do so generally, we should certainly expect recall ability to be related to learning ability. This assumption follows from the fact that retroactive inhibition or interference effects is our principal explanation of loss in ability to recall learned material.

As already noted, some recent work (Stroud & Schoer, 1959; Underwood, 1954) suggests that differences in recall may be unrelated to differences in learning ability. At least this was found to be the case in the experiments in question. At this stage it would seem unwise to generalize very far about this. In the experiments just referred to, subjects learned lists of nonsense syllables, paired adjectives, and picture-names, and presumably went about their normal business during the time between learning and recall. In these respects the materials and general procedure were not different from those employed by earlier workers who (by reason of the erroneous use of relearning as a measure of retention—Stroud & Schoer, 1959) concluded that retentive ability is positively related to learning ability.

In the course of normal events, college students would encounter little material between learning and recall calculated to interfere with the recall of the learned material. At least interference effects should be at a minimum. In any event, it would be interesting to ascertain whether or not differences in learning ability would be found to be associated with differences in recall,

were appropriate interpolated learning systematically introduced. Conceivably retention and learning ability may be related after all.

There is, however, one disquieting thought: a list of nonsense syllables, learned in a psychological laboratory ought to be one of the best remembered things in the world. Hardly anywhere else would one encounter so few things within a 24- or 48-hour retention period to interfere with the recall of the learned material. Obviously, such material is forgotten at a rapid rate. Incidentally, the fact that conventional laboratory material is forgotten so readily under conditions which from the standpoint of interference should be highly favorable to its retention, suggests that there may be factors other than retroactive inhibition operating in forgetting.

Loss in availability of response may in such cases contribute to the relatively rapid rate of forgetting, especially when the subject has only a short interval in which to make a response. Perhaps loss of warm-up effects, or loss of appropriate mental set, which warm-up exercises may help to restore, are involved. Again, this phenomenon may be especially important in the recall of nonsense syllables or pairs of unrelated words, in comparison with the recall of meaningful material, where the subject is frequently able to make use of a rich complement of associative cues.

#### SUMMARY

Intralist inhibition as a function of ability level and length of list has been

investigated. Two groups of subjects widely separated in learning ability learned and recalled (after 48 hours) two lists of paired adjectives of 12 and 24 pairs. All items on both lists by all subjects were learned to a criterion of two and only two correct anticipations. Increasing list length resulted in a disproportional increase in number of trials required to learn, but in a proportional increase in the number of words recalled. Over-all, ability differences in recall were not significant. However, they were significant on the long list. The interspersing of previously learned items, referred to as warm-up items, within the two lists resulted in ability differences in loss of response. The results are discussed in relation to ability differences in recall.

#### REFERENCES

- HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State University of Iowa, 1943.
- ROBINSON, E. S., & DARROW, C. W. Effect of length of list upon memory of numbers. *Amer. J. Psychol.*, 1924, 25, 235-243.
- ROBINSON, E. S., & HERON, W. T. Results of variation in length of memorized material. *J. exp. Psychol.*, 1922, 5, 428-448.
- STROUD, J. B., & SCHOER, L. Individual differences in memory. *J. educ. Psychol.*, 1959, 50, 285-292.
- UNDERWOOD, B. J. Speed of learning and amount retained. *Psychol. Bull.*, 1954, 51, 276-282.

(Received June 20, 1960)

# CONSISTENCY AND WISDOM OF VOCATIONAL PREFERENCE AS INDICES OF VOCATIONAL MATURITY IN THE NINTH GRADE

DONALD E. SUPER

*Teachers College, Columbia University*

The practice of inquiring concerning the vocational preferences and ambitions of junior and senior high school students in helping them to make educational or prevocational choices and plans is widespread and based on compelling arguments: it would be unthinkable, in a democratic society, to require the determination of the directional choices which students make in selecting college preparatory, commercial, trade, home economics, agriculture, and general courses, without taking into account the goal to which the youth aspires. High school guidance questionnaires to be filled out by entering pupils ask what occupation they hope to enter, and numerous studies have been made of ways in which to elicit information concerning vocational preferences (Beilin, 1952; Gilger, 1942; Hamburger, 1958; Trow, 1941).

The consistency, and particularly the wisdom or realism, of vocational preferences have often been used as measures of the effectiveness of vocational guidance programs by practicing counselors and by educational and psychological research workers (Froelich, 1949; Williamson & Bordin, 1941). Having a vocational objective is important in a society in which earning a living is important, in which occupational roles are of major significance, and in which education is, in effect if not avowedly, occupationally oriented: having a vocational preference, in this context, gives purpose to behavior and makes possible educational and vocational decisions. It has been argued that *consistency of vocational*

*preferences* shows intensity and validity of interest, and that it is better to work consistently toward one clear-cut goal than wastefully to keep shifting objectives. In the case of *wisdom of vocational preferences*, the reasoning is that realistic goals are by definition attainable, whereas unrealistic or unwise goals are by definition those which one is not likely to attain or with which one is not likely to be satisfied if one does attain them.

Studies using consistency of vocational preferences are not very common, this being typically a counselor's method of judging how seriously to take a student's expressed preference, but the method is illustrated by Rothney's recent Wisconsin Guidance Study (1958).

Wisdom or realism of vocational preferences has been used as a criterion of the effectiveness of vocational guidance programs in many studies, e.g., Sparling (1933), Kefauver and Hand (1941), Stone (1949), Rothney and Roens (1950), Hoyt (1955), Rothney (1958), and Hewer (1959). Indices of realism are often used by counselors in judging students' and clients' need for guidance.

Despite the widespread acceptance of the importance of consistency and wisdom of vocational preferences, many writers on vocational guidance and on vocational development have questioned the significance of expressed vocational preferences in early adolescence. Fryer's (1931) review pointed clearly to the conclusion, confirmed by Carter's later review (1944), that the expressed preferences of boys and

girls in their early and middle teens are unstable. More recently still, Schmidt and Rothney (1955) reported convincing evidence on the instability of expressed vocational preferences from one year of high school to the next and into the first year out of school: only 49% of the "choices" of the tenth grade remained the same in eleventh grade, and this figure was reduced to 35% in the twelfth grade, and to 24% in the year following graduation.

At the same time, it must be noted that expressed preferences do have practical significance when viewed from a certain vantage point and at certain ages. Dyer (1939) showed that, among college students, vocational preferences which had been constant over a period of years were related to subsequent occupational choice. Strong (1955) and McArthur and Stevens (1955) reported that, among students at Stanford and Harvard, expressed vocational preferences have considerable predictive value for adult occupation. As Dyer's study involved preferences which were constant over a long period of years (omitting all cases in which change of preference had taken place) and the others dealt with expressions of preference in late adolescence and early adulthood among students of superior intellectual, educational, and socioeconomic status, they do not conflict with the studies of more heterogeneous groups of early adolescents which have already been cited.

In view of the persistent use of the concepts of consistency and wisdom of vocational preferences despite their tendency to be unstable in early adolescence, it is important to examine their psychological significance. In the Career Pattern Study of the Horace Mann-Lincoln Institute of School Experimentation several such measures were developed (Super & Overstreet,

1960) and related to other variables in the ninth grade.

## METHOD

### Subjects

The subjects of this study were the core group of the Career Pattern Study of the Horace Mann-Lincoln Institute of School Experimentation, 105 ninth grade boys who were found to be typical in age, intelligence, socioeconomic status, and other key variables of ninth grade boys in Middletown, New York, in the early 1950's. As Middletown itself is an average town on a variety of social and economic indices (Super & Overstreet, 1960), these boys may be considered typical of ninth graders in many American communities. All of the CPS boys indicated, in their interviews, at least one tentative vocational preference.

### Measures of Consistency and Wisdom of Preferences

Three measures of consistency of vocational preferences were developed, the third measure being essentially a combination of the first two. The expressions of vocational preferences were obtained in tape recorded interviews in which the subjects of the study were asked "... about your plans for the future. What would you like to be by the time you're thirty?" This question was followed up with nondirective leads designed to keep the boy talking about his vocational preferences, and with probes designed to make sure that relevant information was obtained if not offered spontaneously (Super, Crites, Hummel, Moser, Overstreet, & Warnath, 1957). In the analysis of the data the first four vocational preferences expressed (if that many were forthcoming) were tabulated. (It should be noted that these are measures of consistency of first and alternative preferences at one point in time, not consistency of first preferences over time.)

An index of *Consistency within Fields* was obtained by classifying the expressed preferences according to the occupational field classification developed by Roe and modified by Moser, Dubin, and Shelsky (1956). The total number of fields into which the boy's preferences fell, minus 1, was his index of *Consistency within Fields*.

A second index, *Consistency within Levels*, was developed by classifying the same preferences according to the modified Roe

occupational level scale, the score being the total number of levels at which expressed preferences fell, minus 1.

A third index, *Consistency within Families*, was developed, by summing the scores on the above two indices, an occupational family being defined as a combination of field and level (e.g., technical occupations at the professional level constitute a family by this definition).

Four measures of wisdom or realism of vocational preferences are relevant to this study. Although no one of these can be viewed as a sufficient index of wisdom of choice, each of them involves a variable which is widely accepted and objectively justifiable as one measure or criterion of realism.

The first index of wisdom or realism was one of *Agreement between Ability and Preference*, or, more accurately, of agreement between the measured intelligence of the individual and the intelligence characteristic of persons employed in the occupation of his first preference. If the boy's score on the Otis Quick-Seoring Mental Ability Test, converted into an AGCT equivalent, exceeded that of the bottom quarter of the men in his preferred occupation as shown by the manual for the Army General Classification Test, his ability was considered to be in agreement with the occupational intelligence requirements; if it was equal to that of men in the bottom quarter, his ability was considered not to be in accord with the requirements.

The second wisdom measure was the index of *Agreement between Measured Interests and Preference*. Interests were measured by the Strong Vocational Interest Blank, slightly modified to insure comprehension at the ninth grade level, and each boy's interests were classified as primary, secondary, or tertiary in the family in which his first preference fell, using Darley's (1941) method of classifying interest score patterns. When the measured interest pattern was primary in the field corresponding to the expressed preference, a score of 4 was assigned, when the pattern was secondary the boy was given a score of 3, etc.

The third wisdom index was that of *Agreement between Occupational Level of Measured Interests and Level of Preference*. The socioeconomic level of the boy's interests (occupational interest level) was measured by the Strong OL scale. If the boy's OL score was not more than one standard deviation below the mean OL score of his preferred occupation, as shown on Table 50

in Strong's monograph (1943, p. 192), the interest and preference levels were considered to be in agreement.

The final wisdom measure to be discussed here is that of the *Socioeconomic Accessibility of Preference*. The family breadwinner's occupation was rated according to the Hamburger revision (1958) of the occupational rating scale included in the Index of Status Characteristics (Warner, Meeker, & Eells, 1949). The boy's vocational preference was rated on the same scale, the agreement or disagreement of the two ratings was ascertained, and the size of the discrepancy was the index of socioeconomic accessibility.

## RESULTS

It should be emphasized that no one of these indices is considered completely satisfactory, but that there are good arguments in support of each of them. For example, the occupational intelligence data of World War II are recognized as an imperfect sample of civilian occupations (Super, 1949), but they are the best available and have proved generally usable; furthermore, the use of the first quartile as the cutting point takes into account the great range of intelligence which makes possible success in any one occupation. In the case of socioeconomic accessibility, it may be objected that it is unwise, in a fluid and democratic society, to judge the realism of a vocational preference by its correspondence with the status of the parents; but even in our democratic society many studies show that the occupation entered by the child tends to be at the same socioeconomic level as that of the parent (Super, 1957), and even when the child changes socioeconomic levels he does it with the help or hindrance of his parents' resources, contacts, information, and values. Each of these measures of the wisdom of a vocational preference is, therefore, one of the possible components of a conceptually more satisfying and valid global index of realism, in which

TABLE 1  
INTERCORRELATIONS OF INDICES OF CONSISTENCY AND WISDOM OF NINTH GRADE  
VOCATIONAL PREFERENCES

Index	Consistency			Wisdom		
	Field	Level	Family	Aptitude	Interest	Level
<b>Consistency:</b>						
Within Fields						
Within Levels	22*					
Within Families	62 <sup>a</sup>	81 <sup>a</sup>				
<b>Wisdom in terms of:</b>						
Aptitudes	16*	06	18*			
Measured Interests	-01	-12	-12	-06		
Measured Interest Level	26**	04	14	00	-32	
Accessibility	08	13	18*	27**	-09	17*

Note.— $N = 105$  boys. Decimal points have been omitted.

<sup>a</sup> Spuriously high because of use of one measure in the other.

\* Significant at the .05 level, one-tailed test.

\*\* Significant at the .01 level, one-tailed test.

having the appropriate intelligence might, for example, offset not having the socioeconomic background which would be a help in achieving one's ambitions.

#### *Agreement among Indices of the Consistency and Wisdom of Ninth Grade Vocational Preferences*

If consistency and wisdom of vocational preferences are valid concepts with which to work in dealing with ninth graders, one should find at least a moderate degree of agreement (correlations of .30 to .50) between various measures of consistency of vocational preferences, a similar degree of agreement among indices of the wisdom of the vocational preference, and some agreement (correlations of .20 to .35) between consistency and wisdom indices. Table 1 reports the intercorrelations of seven CPS indices of consistency and wisdom of vocational preferences, for the 105 ninth grade boys in the core group.

The indices of Consistency of Field and of Consistency of Level are, of course, highly correlated with the index of Consistency of Families: this

happens because the last-named index is a combination of the first two. Only Field and Level in this group of three measures are so constructed as to be operationally independent of each other, so that the correlation between these two measures has meaning; it is .22, significant at more than the .05 level but below the .01. This suggests a slight tendency for the ninth grade boys who are most consistent in aspiring to occupations which are at the same level to be the most consistent in aspiring to occupations which are in the same general field. The implication of this very slight relationship is that the concept of "consistency of vocational preferences" has minimal meaning when boys are in the ninth grade, although it may take on more significance at a later age.

The Consistency indices, Table 1 reveals, are not at all related to Wisdom as measured by the index of Agreement between Interests and Preferences, Wisdom indices. Consistency of the other and Consistency of Field (based partly on Field) are related, to a low and just barely significant degree, to

Agreement of Aptitudes and Preference, and Consistency of Family is similarly related to Accessibility of Preference. The correlation of .26 between Consistency of Field and Agreement between Occupational Level of Measured Interests and Level of Preference is presumably an artifact, as each of these indices is in part based on other measures which deal largely with level: Field involves type of job, and some types of jobs are largely at lower levels (e.g., Business Contact, Outdoor) whereas other types are largely at higher levels (e.g., General Cultural, Science), and Level of Interests or Preferences is by definition and operationally a measure of socioeconomic level. The relationships between Consistency and Wisdom shown by Table 1 may therefore be viewed as negligible.

The internal agreement among the indices of Wisdom of Vocational Preference must next be considered. Agreement between Aptitude and Preference is unrelated to Agreement between Measured Interest and Preference, or to Agreement between Level of Measured Interests and Level of Preference; its correlation of .27 with Accessibility of Preference should probably be viewed as an artifact in view of the facts that the nonpreference variables in the two Agreement measures (intelligence and socioeconomic status) correlate .27 in this sample, and the preference variables (socioeconomic level of preference and intelligence level of preference) in the two measures presumably have an even higher intercorrelation because in each case one variable (preference) is scaled on one of two other highly intercorrelated variables (socioeconomic status and intelligence).

Agreement between Measured Interests and Preference is unrelated to the other Wisdom indices, two of the

intercorrelations being nonsignificant and the third (- .32) being the opposite direction from that hypothesized. And, finally, Agreement between Level of Measured Interests and Level of Preference is also to be viewed as unrelated to the other Wisdom indices, as the barely significant and very low correlation of .17 with Accessibility can be attributed to an artifact: both components of both agreement measures are scaled as to socioeconomic level, and the socioeconomic level of the vocational preference is one of the two measures entering into each of the indices.

The only conclusion that it seems legitimate to draw from Table 1 is, then, that the few seemingly significant intercorrelations may be the products of artifacts, and that the various indices of consistency and wisdom of vocational preferences are unrelated to each other. This lack of relationships suggests a lack of validity as indices of anything significant in the vocational development of ninth grade boys.

#### *Agreement between Consistency and Wisdom of Preferences and Other Variables*

The construct validity of a set of measures of the same variable depends not only upon the intercorrelations of these presumably similar measures, but also upon their agreement with other variables to which theory would lead one to expect them to be related. If consistency and wisdom of vocational preferences are considered to be indices of vocational maturity, of the degree of vocational development which has taken place in an adolescent, then they should be related to measures of other characteristics which might be expected to develop concomitantly with, to result in, or to be the product of, vocational maturity.

TABLE 2  
RELATIONSHIPS OF CONSISTENCY AND WISDOM OF NINTH GRADE VOCATIONAL  
PREFERENCES TO STATUS AND BACKGROUND VARIABLES

	Age in Grade 9	Socioecon. Level	Intell. in Gr. 9	School Achiev.	Pattern Interests	TAT Adj.	Peer Accept.
Consistency:							
Within Fields	-11	-04	-04	.09	-.14	-.04	-.01
Within Levels	-.08	-.03	.06	.00	-.08	.05	.05
Within Families	-12	-02	.02	.00	-.18	.03	-.01
Wisdom in terms of:							
Aptitudes	-16	-.07	.12	.03	-.27	-.10	-.12
Measured Interests	-12	.02	.00	.06	.29**	.14	.07
Measured Interest Level	.14	-.11	-.22	-.06	-.27	-.06	-.17
Accessibility	.12	.42**	-.01	.02	-.09	-.07	-.02

Note.—*N* = 105 boys. Decimal points have been omitted.

\* Significant at the .05 level, one-tailed test.

\*\* Significant at the .01 level, one-tailed test.

The Career Pattern Study therefore used or developed a series of measures of other variables which could be hypothesized as related to vocational maturity. These were age (except that the fact that the negative correlation between age and intelligence in any one grade—in which the older pupils tend to be the retarded and the younger tend to be the accelerated—and the limited age range may be expected to confuse the relationship), socioeconomic level, intelligence, school achievement, patterning of vocational interests, emotional adjustment, and peer acceptance.

Socioeconomic level was measured by the placement of the family breadwinner's occupation on the Hamburger revision of the Warner scale, intelligence by the Otis Quick-Scoring Test of Mental Ability, school achievement by grades in the three constant courses (taken by all students) of the ninth grade, patterning of interests by the application of Darley's (1941) method to scores on Strong's Vocational Interest Blank, adjustment by Overstreet's method of deriving a total adjustment score from stories told in the Thematic Apperception Test (Super & Overstreet, 1960), and peer

acceptance by a modification of the Guess Who technique (Super & Overstreet, 1960).

The correlations of the Consistency and Wisdom indices with these variables are reported in Table 2. Only 2 of the 49 correlations are statistically significant and in the hypothesized direction, and one of these relationships is due to an artifact: Accessibility is correlated .42 with Socioeconomic Level, but the latter constitutes part of the former. The one presumably true relationship is that of Agreement between Measured Interests and Preferences with Patterning of Interests (.29): boys whose preferences agree with their measured interests tend to have clear-cut patterns of measured interests, a relationship which makes excellent psychological sense and which gives one a little confidence in measuring interests as having some meaning at the ninth grade level, despite their lack of relationship to other variables. (But when interests are related only to interests one wonders if there is perhaps a measurement artifact at work here also.)

Five other correlations in Table 2 are large enough to be significant, but since they are in the unexpected

direction and do not make psychological sense they must be attributed to chance. In a table of 49 correlations, one might find 2 or 3 correlations which seem statistically significant at the .05 level, and 1 correlation significant at the .01 level, strictly on a chance basis.

The conclusion to be drawn from Table 2, as from Table 1, thus appears to be that the consistency and wisdom of vocational preferences have little significance for prevocational and vocational choices at the ninth grade level.

### *Implications*

If the consistency and wisdom or realism of a ninth grade boy's expressed vocational preferences lack validity, as the Career Pattern Study data suggest, does this mean that they should be disregarded in practice? The fact that consistency and wisdom of vocational preferences have little meaning at Grade 9 does suggest that they should not be used as criteria of the need for guidance nor of the effectiveness of guidance at that stage of development. Presumably at this stage the very instability and transiency of expressed vocational preferences, and perhaps also their inadequate factual basis, make their apparent consistency and wisdom largely a matter of chance. However, one would be reluctant to draw the conclusion that the preferences themselves, even if inconsistent or unwise, should be disregarded, for this would involve not only the assumption that the counselor knows better than the pupil what is appropriate for him, but also the assumption that the best way to help the pupil to choose and plan wisely is to get him to concentrate on data coming from without himself rather than to examine his self-concept in relation to external

and impersonal data. The demonstrated, even though far from perfect, validity of aptitude, achievement, and interest test data make it clear that the counselor does indeed have unique externally derived information as to what is appropriate for the pupil. But it may well be, as many counselors believe and as some research shows, that the best way to let this information help the pupil is to aid him in assimilating it into his concept of himself. This is best done by beginning, not with the data, but with the self-concept. And the statement of a vocational preference is one way of expressing a self-concept, as Bordin (1943) and the present writer (1951) have pointed out.

There are, of course, various ways of helping in the exploration of the self-concept as manifested in an expressed vocational preference. One is to help the pupil to choose courses, activities, part-time employment, etc. in which he may find opportunities to see if the preferred role does indeed suit him, and to discuss his use and evaluation of these experiences with him as he participates in them. Another is to help him to examine his preference in the interview, relating it to his picture of himself as a student, as a part-time worker, as a member of a family group, etc. In this discussion he may be helped to consider the views of him held by other persons and the picture of himself obtainable from records of his performance. This discussion can lead to the desirability of obtaining other such external data, and more testing or more participating in courses or activities may be planned and the results later reviewed by pupil and counselor. By thus using expressed vocational preferences as springboards for exploration and growth, for self-evaluation and further planning, the counselor proceeds democratically

and with respect for his client's right to and need for self-determination, while avoiding the error of placing undue emphasis on preferences as a basis for directional choice.

### SUMMARY

Questions are raised concerning the significance of the consistency and wisdom or realism of vocational preferences among ninth grade boys, and the use of measures of these as criteria of the effectiveness of or the need for vocational guidance. The development and application of several such measures in the Career Pattern Study are described, and data on the construct validity of these indices for 105 typical boys are reported. The failure to find significant relationships in the hypothesized directions is taken as evidence of the lack of psychological and hence of practical educational significance of consistency and wisdom or realism of vocational preferences at this stage of development. It is concluded that, although they may be meaningful at later stages, they should not be used as criteria of the need for or effectiveness of guidance and counseling at the ninth grade level.

### REFERENCES

- BEILIN, H. Factors affecting occupational choice in a lower socio-economic group. Unpublished doctoral dissertation, Teachers College, Columbia University, 1952.
- BORDIN, E. S. A theory of vocational interests as dynamic phenomena. *Educ. psychol. Measmt.*, 1943, 3, 49-65.
- CARTER, H. D. Vocational interests and job orientation. *Appl. psychol. Monogr.*, 1944, No. 2.
- DARLEY, J. G. *Clinical application and interpretation of the Strong Vocational Interest Blank*. New York: Psychological Corp., 1941.
- DYER, DOROTHY T. Relation between vocational interests of men in college and their subsequent histories for ten years. *J. appl. Psychol.*, 1939, 23, 280-288.
- FROELICH, C. P. *Evaluating guidance procedures: A review of the literature*. Washington, D. C.: Office of Education, 1949.
- FRYER, D. *The measurement of interests*. New York: Holt, 1931.
- GILGER, G. A., JR. Declaration of vocational interest. *Occupations*, 1942, 20, 276-279.
- HAMBURGER, M. Realism and consistency in early adolescent aspirations and expectations. Unpublished doctoral dissertation, Teachers College, Columbia University, 1958.
- HEWER, VIVIAN H. Group counseling, individual counseling, and a college class in vocations. *Personnel guid. J.*, 1959, 37, 660-665.
- HOYT, D. P. An evaluation of group and individual programs in vocational guidance. *J. appl. Psychol.*, 1955, 39, 26-30.
- KEFAUVER, G. N., & HAND, H. C. *Appraising guidance in secondary schools*. New York: Macmillan, 1941.
- MCAULIFFE, C., & STEVENS, LUCIA B. The validation of expressed interests as compared with inventoried interests: A fourteen-year follow-up. *J. appl. Psychol.*, 1955, 39, 184-189.
- MOSER, HELEN P., DUBIN, W., & SHELSKY, I. A proposed modification of the Roe Occupational Classification. *J. counsel. Psychol.*, 1956, 3, 27-31.
- ROTHNEY, J. W. M. *Guidance practices and results*. New York: Harper, 1958.
- ROTHNEY, J. W. M., & ROENS, B. A. *Guidance of American youth*. Cambridge, Mass.: Harvard Univer. Press, 1950.
- SCHMIDT, J. L., & ROTHNEY, J. W. M. Variability of vocational choices of high school students. *Personnel guid. J.*, 1955, 34, 142-146.
- SPARLING, E. J. Do college students choose vocations wisely? *Teach. Coll. Contrib. Educ.*, 1933, No. 561.
- STONE, C. H. Are vocational orientation courses worth their salt? *Educ. psychol. Measmt.*, 1949, 8, 161-181.
- STRONG, E. K., JR. *Vocational interests of men and women*. Stanford: Stanford Univer. Press, 1943. P. 192.
- STRONG, E. K., JR. *Vocational interests 18 years after college*. Minneapolis: Minnesota Press, 1955.
- SUPER, D. E. *Appraising vocational fitness by means of psychological tests*. New York: Harper, 1949.
- SUPER, D. E. Vocational adjustment: Implementing a self-concept. *Occupations*, 1951, 30, 88-92.

- SUPER, D. E. *The psychology of careers*. New York: Harper, 1957.
- SUPER, D. E., CRITES, J. O., HUMMEL, R. C., MOSER, HELEN P., OVERSTREET, PHOEBE L., & WARNATH, C. F. *Vocational development: A framework for research*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1957.
- SUPER, D. E., & OVERSTREET, PHOEBE L. *The vocational maturity of ninth-grade boys*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1960.
- TROW, W. D. Phantasy and vocational choice. *Occupations*, 1941, 20, 89-93.
- WARNER, W. L., MEEKER, MARCHIA, & EELS, K. *Social class in America*. Chicago: Science Research Associates, 1949.
- WILLIAMSON, E. G., & BORDIN, E. S. The evaluation of vocational and educational counseling. *Educ. psychol. Measmt.*, 1941, 1, 5-24.

(Received June 20, 1960)

## THE COMPARATIVE INFLUENCE OF PUNITIVE AND NONPUNITIVE TEACHERS UPON CHILDRENS' CONCEPTS OF SCHOOL MISCONDUCT<sup>1</sup>

JACOB S. KOUNIN AND PAUL V. GUMP

*Wayne State University*

This paper reports a portion of a research project pertaining to the management of childrens' behavior in classroom settings. Because so many teachers, especially beginners, verbalize considerable concern about discipline and control, we are focusing our current research in this area. While there is some relevant literature, such as that of Sheviakov and Redl (1944), based upon experience and insightfulness, we have been unable to locate any generalizations based upon data from research.

In a previous study by Kounin and Gump (1958) specimen-record types of observations were gathered of discipline incidents during the first week of kindergarten, focusing upon the triad of: a misbehaving child (target), a teacher doing something to stop the misbehavior, and a watching audience-child. Limiting our dependent variables to overt behavior we found that teachers' techniques of handling a misbehaving kindergarten child (target) did have different degrees of socializing success upon audience-children. A socializing success was defined as an observable reduction of overt misbehavior or an increase in conforming behavior (standing up "even straighter" in line). Control techniques high in clarity (defining the deviancy, specifying

ing how to stop) were most successful. Control techniques high in firmness (standing closer to the misbehaving child, continuing to look at him until he stopped misbehaving) were successful only for audience-children who were themselves deviancy-oriented at the time. Control techniques high in roughness (anger, physical handling) were least successful and tended to be followed by behavior disruption (less involvement in work, overt signs of anxiety) rather than conformity on the part of audience-children. In terms of their effects, it is evident that roughness is a different dimension than firmness.

Since attitudes toward misconduct may also be affected by differences in control techniques we decided to study these as well. In an unpublished study of children at camp, P. Gump, B. Biddle, and J. Kounin found significant differences in attitudes toward camp misconduct held by campers who had effective counsellors as compared to campers who had ineffective counsellors. The counsellors, however, varied along many dimensions including punitiveness, goal-directedness, physical and psychological absenteeism, and misconduct also varied according to whether they were talking about camp, home, or school milieus. We decided, therefore, to limit the leadership dimension to punitiveness and the milieu to school.

It is postulated that aggression leads to counteraggression; it is further postulated that a punitive teacher has more power over her pupils than they have

<sup>1</sup> A version of this paper was presented at the American Psychological Association meeting, September 5, 1959. It is part of an investigation supported in part by a Research Grant M-1066, from the National Institute of Mental Health, United States Public Health Service and in part by a grant from the College of Education of Wayne State University.

over her and that she blocks overt manifestation of pupils' aggression (observations in the classrooms of the punitive teachers selected for this study indicate that this second assumption is tenable). From these two postulations, we derive the following hypotheses:

1. That the school misconduct pre-occupations of children with punitive teachers will contain more aggression than those of children with nonpunitive teachers.
2. That children with punitive teachers will be more conflicted about school misconduct than will children with nonpunitive teachers.
3. That the aggression needs and the conflict relating to misconduct hypothesized to exist among children with punitive teachers will detract from their concern with school-unique values that are not directly related to misconduct.
4. The question may also be raised as to whether or not the amount of tension generated in the children with these particular punitive teachers is sufficiently great to reduce the rational qualities of their attitudes toward misconduct.

### METHOD

*Subjects.* The subjects were 74 boys and 100 girls attending their first semester of the first grade in the public schools of a large city. They represented all the children from six home rooms of three schools, in from upper-lower to middle-middle socio-economic neighborhoods.

*Procedure.* Overall school climate was controlled by selecting pairs of punitive vs. nonpunitive teachers from the same school. Three such pairs were obtained from three elementary schools.

The initial selection of punitive and non-punitive teachers was obtained from the principal and assistant principal. Following this the classes were observed by both principal investigators. At approximately a week later the teachers were further rated by a supervisor of student teachers who visited each class twice.

The raters checked along a continuum from Extremely Punitive (threatens children with consequences that really hurt; makes threats that imply sharp dislike, real willingness to harm child; ever-readiness to punish) to Not Punitive (does not punish and does not threaten). A punitive vs. non-punitive pair of teachers was used for the study only when all five persons agreed on their dichotomizations. All the teachers were rated as having good organization, well-behaved classes, and as achieving the learning objectives for their grade. Eighty-four of the children were in classes with punitive teachers and 90 children were in classes with nonpunitive teachers.

The children were interviewed individually during the third month of attendance at school. The interview consisted of the questions: "What is the worst thing a child can do at school?" and, following the reply, "Why is that so bad?" Identical questions were asked regarding home as the milieu for misbehavior.

*Coding the Replies.* The misconducts mentioned by the children were coded for content and for certain qualities or dimensions.

The content code (obtained from the question of "What is the worst thing to do?") contained two parts: the misconducts and the explanations given for why these were bad. The misconduct included: the act type (physical or psychological assaults, noncompliance, etc.) and the object of the misconduct (parents or teachers, other children, institutional laws or custom, etc.).

The code for the explanation of misconduct was designed to answer three questions: Who is involved in the consequence (the child himself, parent or teacher, a peer, etc.)? What kind of sufferings result to others from the misconduct (physical pain, achievement loss, property loss, etc.)? What kinds of retributions occur to the misbehaver (work imposals, character loss, physical punishment, etc.)?

### RESULTS AND CONCLUSIONS

Children probably answer the question of "What's the worst thing a child can do in school?" with a report of acts that reflect their preoccupations. It is not likely that our subjects' answers would have been the same if they were presented with a forced-choice of alternative acts. Given a

choice, most children would probably rate "stabbing someone" as more serious than "talking in class." If the misconducts the children talked about are taken to represent tension systems and preoccupations, we may infer from these the comparative impact of punitive and nonpunitive teachers.

In a concurrent study of childrens' attitudes toward misconduct (201 boys and 214 girls in the first grade of six public schools representing a range of socioeconomic backgrounds), Gump and Kounin (1959) found both sex differences and differences between home and school milieus. For example, home misconducts included more breaking of objects while school misconducts included more rule violations; parents suffered more than teachers in consequences but teachers retributed more frequently. However, parents were reported as retributing with more corporal punishment and with more severe punishment than teachers. There were also differences in the responses of boys and girls, especially in school. For example, girls reported "talking" as a school misconduct eight times more frequently than boys, whereas boys reported physical assaults on peers in school more frequently than did girls.

Consequently, the comparison of the responses of children with punitive and nonpunitive teachers was made separately for sexes and also for home and school milieus. However, on all comparisons of school responses the direction of differences between children with punitive and nonpunitive teachers was the same for both boys and girls. The report of results, therefore, combines both boys and girls. Insofar as the differences between children with punitive and nonpunitive teachers are concerned, only 2 of the 48 comparisons of home responses were statistically significant: home miscon-

ducts of the children with punitive teachers were rated as more serious ( $p < .05$ , for girls only) and retributions to the subject were more serious ( $p < .02$ , for boys only). It is uncertain whether these represent some spillover of the influence of punitive teachers onto attitudes toward home misconducts, or whether they are chance differences for the number of comparisons made.

The results to be reported here, then, refer to boys and girls combined and to school misconducts only. Intercoder reliabilities ranged from 73-95% agreement, with a median of 90. The  $p$  levels of differences are based on the  $\chi^2$  test. In the case of dimensions, such as "seriousness," the results were dichotomized into a High and a Low based upon as equal a break as was possible and resulting  $\chi^2$ 's were based upon  $2 \times 2$  tables. In the case of categories such as act-types falling into the categories of: rule violations, physical assaults on children, property damages, or nonconformance with adults, the  $\chi^2$  was computed for as many cells as there were categories. At times, when one particular category was of interest, a  $2 \times 2$  table was constructed with that category versus "all the rest," providing that the overall table showed statistical significance.

Following are definitions of the codes used which are not self-explanatory both for the misconducts and for the subjects' explanations for the wrongness of the act. These codes appear in Table 1.

## I.

- A. Physical assaults include all physical attacks on other persons (pushing, hitting).
- B. Milieu-seriousness refers to the length to which the milieu would go to prevent such an act. The school would practically ignore "scratching head," would mildly frown at "whispering," and would go to any length to stop burning down buildings.
- C. Coder seriousness refers to the general

immorality or danger in the misconduct considered from the point of view of the overall culture. (The only frame of reference which produced high intercoder agreement was when the coders took the position of an understanding Reformed rabbi or a Unitarian minister. Taking an unspecified role, or that of either a parent or a teacher produced low intercoder agreement.) "Studying spelling lessons at the wrong time" is morally trivial while "maiming someone" is morally very serious.

D. Abstractness of misconduct refers to the size of coverage. It may range from a unique, "one time" misbehavior, such as, "cut a climbing rope in gym," to an abstract one, such as, "be mean to other people."

## II.

A. A central adult is the responsible leader: teacher at school.

B. A psychological loss to another is exemplified by "It would make her worry."

C. Seriousness of consequences to others range from trivial harm, such as, "She'd be annoyed" to serious ones, such as, "He'd die."

D. A reality-centered retribution (this is scored only when the perpetrator himself suffers in the consequence) is coded when the consequence of a misconduct follows naturally from the act-type, such as: "not study because you'll get behind in your work." This contrasts with the response in which the connection between act and consequence is dependent upon a personal intervention of another, such as: "not study because teacher will make you stand in the corner."

E. "Reflexive justification" was coded when the child gave no consequence for either himself or others in his explanation of why the act was bad. When he said the act is bad because "It's not nice" or "It's bad," it was called a reflexive justification.

## III.

A. On ego-acceptability, we sought to determine the degree to which the respondent could see himself as the perpetrator of the misconduct. In an ego-alien act, the respondent expresses abhorrence, such as: "It's dirty to hit little kids who didn't do nothing to you." An ego-attractive act is one in which the child indicates its seductive quality for him, such as: "Tell off a teacher —boy, I'd like to do that."

B. On the premeditation category, we sought to learn the extent to which the

TABLE 1  
A COMPARISON OF ATTITUDES TOWARD  
SCHOOL MISCONDUCTS HELD BY CHILDREN  
WITH PUNITIVE AND NONPUNITIVE  
FIRST GRADE TEACHERS  
(N = 176)

Misconducts and Explanations	% of Pu <sup>a</sup>	% of NPu <sup>a</sup>
I. Content and quality of the misconducts:		
A. Physical assaults on others	38	17
B. Milieu-serious misconducts	89	63
C. Coder-serious misconducts	48	27
D. Abstract misconducts	27	52
II. Content and quality of the explanations:		
A. Peers as objects of consequences	94	61
B. Physical damage to objects of consequences	60	23
C. Serious harm to others	45	18
D. Reality-centered retributions	21	48
E. "Reflexive justifications" as explanations	11	26
III. Role of self in misconducts:		
A. Ego-alien misconducts	26	11
B. Premeditated misconducts	29	15
IV. Aggression:		
A. Overall aggression ("blood and guts")	49	24
V. Concern with school-unique objectives:		
A. Learning and achievement losses	20	43
B. Institutional law violations	49	62

Note.—All differences in percentages are significant at the .05 level or beyond.

<sup>a</sup> Pu stands for those children who have punitive teachers; NPu refers to those children who have non-punitive teachers.

child sought to do wrong. If premeditated, the child plans the act and intends the consequences ahead of time, such as: "Put thumb tacks on teacher's chair when she is out." If intentional, the child accepts his part in the wrongdoing but does not plan it, such as, "talk during a lesson."

## IV.

Aggression ("blood and guts") refers to the amount of aggression the respondent

expresses in his misconducts and consequences. "Play in the storage bin because somebody might get hurt" expresses less aggression than "Play in the storage bin because you might push a kid off and there could be a sharp rock down there and he could hit his head against it and crash open his skull and he would bleed and his brains would fall out and he'd die."

#### V.

A. A learning or achievement loss is coded when interference with learning is the misconduct or the explanation, such as: "It's bad to make noise because somebody could make a mistake in his work," or "... because then he couldn't read good."

B. An institutional law violation is a violation of the rules of the school such as: "talk when you're supposed to study," "not take your seat when the bell rings."

The results presented in Table 1 may be summarized around the three hypotheses and the one question raised in the introduction:

*Punitive teachers will create or activate more aggression-tension than will non-punitive teachers.* This is strongly supported by the data. The children who have punitive teachers have more sheer aggression in their sins and consequences, they give both more milieu-serious and more coder-serious misconducts, their targets suffer more harm, they give more physical assaults as act-types, and their targets suffer more physical harm. The targets of children with nonpunitive teachers are more inclined to suffer psychological losses as consequences. As an example of the results: of 84 respondents with punitive teachers, 31 give physical assaults on other children and 40 mention school rule violations; while of 90 children with nonpunitive teachers, 15 talk about physical assaults and 56 about rule violations. (The remainder of the act-types are nonconformances and "miscellaneous.")

*Children with punitive teachers will be more unsettled and conflicted about misbehavior in school.* This hypothesis

is supported by the findings related to the role of self in misbehavior. The children from nonpunitive teachers give misconducts in which their own role is intentional whereas children from punitive teachers give both pre-meditated and ego-alien misconducts. We may say that children with punitive teachers express more abhorrence for the misdeeds which they have selected and yet select misdeeds which require "malice and forethought."

*Punitiveness of teachers will detract from childrens' concern with school-unique values.* This hypothesis is supported. Children from punitive teachers talk more about physical attacks on peers—misbehavior by no means unique to the classroom setting. Children with nonpunitive teachers talk more about learning, achievement losses, and violations of school-unique values and rules.

*Do children from nonpunitive teachers show more rational qualities in their responses?* The answer to this question is not clear. Fairly direct attempts to measure this—codes for milieu likelihood of misconducts, for likelihood of consequences, and for appropriateness of consequences to the misconduct—did not show significant differences between the two groups. On the other hand, children with punitive teachers gave fewer abstract misconducts which result, in our camp study, was negatively correlated with age. But these same children also gave fewer reflexive justifications which result was positive. One interpretation of the findings that children of punitive teachers gave fewer abstract misconducts and fewer reflexive justifications is to regard these as indications of the unsettled and conflicted state of the attitudes regarding misconduct held by children with punitive teachers. When a child is inclined to misbehave but fears to, then

a concrete act occurs to him—"hit George in the mouth"; when he is not pressed by his needs to misbehave, then an abstraction occurs to him—"be mean to people." Similarly, when he expresses this verbal act, a real consequence occurs—the target gets hurt or the perpetrator suffers a consequence; when he is not preoccupied with wrongdoing then a reflexive justification occurs to him—"it's not nice." A reflexive justification at this age may not be a primitive reply but a reflection of a settled issue: "You just don't do this because it's not nice."

Another interpretation is to regard the greater use of reflexive justification by the children with nonpunitive teachers as evidence of their greater trust and faith in school, i.e., of their internalization of school values more than children with punitive teachers. Inspection of the data showed the reflexive justification was used predominately in connection with rule violations (talking, running in halls, not taking seat, and the like). These misconducts are milieu-inconvenient which are disturbing to the milieu but which are without direct harm to either the actor or to others and do not violate an important moral code. Such misconducts to the first grade child have no real explanation except that "they're bad because they say so." As such, they express a sort of naive faith and trust in the rightness of what the teacher says.

## SUMMARY

Three pairs of punitive vs. nonpunitive first grade teachers were selected from three elementary schools. The 174 children in these teachers' classrooms were individually interviewed about what they thought was "the worst thing to do in school" together with their explanations of why these misconducts were bad. Regarding their responses as expressions of their pre-occupations it was concluded that, as compared with children who have non-punitive teachers, children who have punitive teachers: manifest more aggression in their misconducts, are more unsettled and conflicted about misconduct in school, are less concerned with learning and school-unique values, show some, but not consistent, indication of a reduction in rationality pertaining to school misconduct. A theory that children with punitive teachers develop less trust of school than do children with nonpunitive teachers was also presented to explain some of the findings.

## REFERENCES

- GUMP, P. V., & KOUNIN, J. S. Sex and milieu influences in children's concepts of misconduct. *Amer. Psychologist*, 1959, 14, 346. (Abstract)
- KOUNIN, J. S., & GUMP, P. V. The ripple effect in discipline. *Elem. sch. J.*, 1958, 59, 158-162.
- SHEVIAKOV, G., & REDL, F. Discipline for today's children and youth. Washington, D. C.: National Education Association, 1944.

(Received June 29, 1960)

# A LONGITUDINAL STUDY OF THE PRIMARY MENTAL ABILITIES TEST

WILLIAM J. MEYER AND A. W. BENDIG

*University of Pittsburgh*

A major criticism of research concerned with developmental changes in the structure of intelligence has been the lack, at each age level, of comparable test batteries and subjects (Anastasi, 1948). It can be argued that inappropriate test content at one age level or another serves to lower reliability which in itself could account for age changes. Differences in variability of scores can similarly affect the usual statistical procedures employed in these kinds of studies. Accordingly, longitudinal data would seem to be one means of reducing the effects of these variables. The writers have fortunately been able to employ a sample of 100 subjects who were examined on two occasions separated by a time interval of approximately 3.5 years with a test standardized for both developmental levels. The purpose of this paper is to present several analyses of the data appropriate to each of the following issues: developmental changes in the magnitude of *g*, consistency of relative position and consistency of profiles, consistency of factorial structure (factor validity), developmental changes in magnitude of sex differences and rates of growth on each of the primary abilities, and accuracy of long range predictions of achievement.

## METHOD

The Primary Mental Abilities Test (PMA), Intermediate Form (age 11-17), was administered to all of the eighth graders present in a junior high school located in a small industrial community.<sup>1</sup> Approx-

mately 3.5 years later, during the last week of Grade 11, the same form of the test was readministered, at which time only 8% of the original sample were, for a variety of reasons, not in attendance. As shown in Table 1, the final sample of 100 youngsters, 49 boys and 51 girls, scored somewhat above average on the test ( $M_{IQ} = 107.95$ ) with the variability comparing favorably with the standardization group ( $\sigma = 16.56$ ). As a safeguard against selective factors influencing the data, only those subjects available at both testing sessions were included in the final analysis. An important feature of this sample is the fact that the subjects are homogeneous with respect to socioeconomic background (lower middle class) and education. Most of the boys were in the general education program whereas most of the girls were enrolled in a commercial program.

The PMA yields scores on the five presumably independent traits which Thurstone (1938) has defined as constituting intelligence. These five traits include Verbal Meaning (V), Space (S), Reasoning (R), Numerical (N), and Word Fluency (W). In addition a Total score (T) is available. The reliabilities of each of the subtests as well as the Total score are quite satisfactory (Thurstone, 1958). It would have been helpful to have available reliabilities by age or grade levels, but unfortunately such data are apparently unavailable in published form. Since the test was standardized for an age group inclusive of the present sample, the assumption is made that reliabilities are essentially comparable over the age range inclusive of Grades 8 and 11. In addition to the foregoing psychometric features, this scale was employed because the types of analysis planned had not been previously reported in the literature for the PMA.

On the day following the administration of the PMA, the Myers-Ruch High School Achievement Test (MRT) was administered. Eight students were absent for this test so that analyses including the MRT are

<sup>1</sup> Appreciation is expressed to H. S. Konvolinka, Superintendent of Schools,

Duquesne, Pennsylvania, and to H. McKeegan, guidance counselor at the high school, for their cooperation in the conduct of this study.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF  
CHRONOLOGICAL AGE AND IQ AT GRADE  
8 AND GRADE 11

Statistic	Grade 8			Grade 11		
	B	G	T	B	G	T
N	49	51	100	49	51	100
$M_{ca}$	12.84	12.72	12.78	16.28	16.17	16.23
$\sigma_{ca}$	.67	.43	.57	.62	.42	.55
$M_{IQ}$	107.69	108.20	107.95	106.24	111.65	108.51
$\sigma_{IQ}$	17.78	14.41	16.56	15.81	14.30	15.05

based on an  $N$  of 92, 46 boys and 46 girls. The MRT was selected because of the short testing time (one hour) and the fact that it samples a variety of subject matter areas. It would probably have been more desirable to use one of the achievement tests yielding a variety of specific scores rather than an overall score, as does the MRT, but the time necessary to administer such tests was not available.

## RESULTS

### *Developmental Changes in the Magnitude of g*

Garrett (1946) in his well known developmental theory of intelligence suggests that the relative dominance of Spearman's  $g$  decreases with age, resulting in the emergence of specific factors. The inconsistent findings reported in the literature would seem to be the result of the methodological problems inherent in the cross-sectional approach (Anastasi, 1948). A study by Asch (1936), who retested in Grade 7 the subjects originally tested by Schiller (1934) in the third grade, deserves closer attention because of the longitudinal approach employed. Consistent with Garrett's position, lower correlations between the verbal and numerical tests were found after the 4 years. It should be noted, however, that only 161 of the original sample of 395 were available for the

second testing. In addition to the effects of losing 59% of his original sample, the degree to which the tests were appropriate for the older subjects was questionable.

The present data were analyzed in a manner consistent with earlier studies so as to provide a basis for comparison. Several investigators (Asch, 1936; Garrett, Bryan, & Perl, 1935) have reported median  $r$ 's derived from intercorrelations among the subtests at each age level. A similar analysis presented in Table 2 for the total group ( $N = 100$ ) reveals a slight decrease (.02) in the magnitude of the intercorrelations. In a separate analysis by sex, also shown in Table 2, the median intercorrelations for boys were .31 and .28 and for the girls they were .32 and .27, indicating little or no sex difference. A second analysis, involving the determination of the proportion of variance accounted for in the first unrotated centroid factor, supports the foregoing conclusions (the percentages being 39% and 37%, respectively). Since both analyses are assumed to be related to the magnitude of  $g$ , it may be concluded that within the age range of the sample there is little or no evidence for increased differentiation of abilities. It should be noted, however, that several of the intercorrelations are moderately high, particularly those involving the V subtest, suggesting that the  $g$  factor is present.

Previous research (Meyer, 1960; Thurstone, 1938; Thurstone & Thurstone, 1941; Tyler, 1958) along with the present data include a developmental span ranging from Grade 1 through senior high school and up to the age of 25 permitting tentative conclusions concerning age-grade changes in performance on the PMA. Tyler (1958) reports a longitudinal study in which the performance of the youngsters in the first grade is com-

TABLE 2  
INTERCORRELATIONS AMONG PRIMARY MENTAL ABILITIES SUBSCALES AT GRADE 8  
AND AT GRADE 11

Group	PMA Subscales						Mean Correlation	
		V	S	R	N	W	Grade 8	Grade 11
Boys (N = 49)	V		.38	.52	.56	.47	.31	.28
	S	.37		.29	.17	.06		
	R	.43	.12		.43	.03		
	N	.53	.11	.56		.19		
	W	.26	.10	.17	.18			
Girls (N = 51)	V		.41	.30	.49	.26	.32	.27
	S	.38		.42	.16	.19		
	R	.42	.31		.30	.39		
	N	.54	.07	.48		.24		
	W	.12	.29	.12	.04			
Total (N = 100)	V		.37	.42	.53	.38	.31	.29
	S	.34		.33	.14	.10		
	R	.46	.18		.38	.20		
	N	.55	.06	.54		.24		
	W	.24	.15	.18	.16			

Note.—Grade 8, above principal diagonal; Grade 11, below principal diagonal.

pared with their performance when in the fourth grade which in turn is compared with their eighth grade scores. From Grade 1 to Grade 4, it was found that total score predicted subsequent performance on each subtest about as well as prior performance on the specific subtest itself. The comparison from Grade 4 to Grade 8 revealed that for N the correlation between corresponding subscores was considerably higher than the correlation between Grade 4 T score and Grade 8 N score. Meyer (1960), who performed the same analysis on the PMA from Grade 8 to Grade 11, found that in addition to N the S subtest also emerges. It should be noted, however, that all of the corresponding subtest correlations were higher than the correlations between total score and each subsequent sub-score. Thurstone (1938) working with eighth grade children reports a correlation between the six primary factors (which he called a second-order general

factor), but in another study (Thurstone & Thurstone, 1941) in which the age range was from 16 to 25, no such second-order factor emerged. From these studies, then, the conclusion may be made that in the primary grades (1-4) the PMA abilities measure general ability whereas from Grade 4 to 8 numerical ability emerges from the others as a more discrete aptitude. After Grade 8, a spatial ability emerges, but not until the Age Group 16 to 25 is there evidence for independent factors.

#### *Consistency of Relative Position and Profile Differences*

An important characteristic of any test is the degree to which present performance predicts future ratings on the particular variable. Certainly many things can happen to an individual which might effect a significant change in his performance relative to his group which would serve to minimize the

importance of prior information. Doppelt and Bennett (1951) examined the long range consistency of the Differential Aptitude Tests (DAT), reporting rather high correlations over a period of 3 years with very little differences between the sexes. Certain tests, such as Verbal Reasoning, and Numerical Ability, showed more consistency over time than tests such as Abstract Reasoning, Space Relations, and Clerical Ability, prompting the speculation that such differences in consistency are related to a uniformity of experiences in the former abilities whereas the latter, being nonschool subjects, are more subject to individual interests and experiences. The present data are relevant to this hypothesis.

Before considering the data analysis procedures followed in this section, it will be necessary to examine the reliability of the subtests. Obviously if the immediate reliabilities of the subscales are low, consistency measures over time would be meaningless. The following split-half coefficients have been reported in the manual (Thurstone, 1958) for tenth graders: V, .92; S, .96; R, .93; N, .89; and W, .72.<sup>2</sup> These reliability coefficients are sufficiently high so that any observed changes in relative position over the 3.5 years could reasonably be ascribed to factors other than errors of measurement.

A summary of the analysis is presented in Table 3. These test-retest correlations are based on the same form of the PMA and, for the total group, are remarkably high except for W. The fact that performance on V, R, and N are more stable over time than are S

TABLE 3  
CORRELATIONS BETWEEN GRADE 8 AND  
GRADE 11 SCORES ON THE PRIMARY  
MENTAL ABILITIES TEST

	Boys	Girls	Total
Verbal Meaning	.88 (.96)	.74 (.80)	.81 (.88)
Space	.51 (.84)	.77 (.80)	.66 (.69)
Reasoning	.76 (.81)	.71 (.76)	.75 (.81)
Number	.83 (.93)	.63 (.71)	.73 (.82)
Word Fluency	.52 (.72)	.27 (.38)	.43 (.59)
N	49	51	100

Note.—Entries in parentheses are coefficients corrected for attenuation.

and W is consistent with the findings of Doppelt and Bennett (1951) described above and would seem to support the notion that the academic abilities are more stable than the non-academic. In contrast with previous findings, our data suggest the existence of sex differences in consistency. Statistical tests support this observation for V, S, and N ( $z = 2.08, 2.20, 2.17$ , respectively), but not for W and R ( $z = 1.45$  and  $.52$ , respectively). These data can be interpreted as meaning that over the Grade Range 8-11 the relative position of boys as contrasted with girls is generally more stable.

Another problem of particular importance when dealing with specific traits as on the PMA is that of profile stability. Guidance counselors in particular are concerned with the long term stability of difference scores since they are often used to determine a person's future program of study. This problem has also been examined by Doppelt and Bennett (1951), for the DAT, who report a median correlation of .50 for all possible combinations of differences between test scores with a range from .20 (Numerical Ability minus Space Relations) to .74 (Mechanical Reasoning minus Spelling). Little in the way of sex differences was noted.

<sup>2</sup> Reliability for the Word Fluency subtest was determined by use of the separately-timed halves technique originally reported by Anastasi and Drake (1954).

TABLE 4  
RETEST CORRELATIONS BETWEEN PRIMARY  
MENTAL ABILITIES SUBTEST DIFFERENCE  
SCORES AT GRADES 8 AND 11

PMA Subtests	Boys (N = 49)	Girls (N = 51)	Total (N = 100)
V - S	.30	-.12	.14
V - R	.65	.19	.42
V - N	.68	.10	.39
V - W	.56	-.02	.27
S - R	.53	.07	.37
S - N	.66	-.10	.31
S - W	.44	.06	.33
R - N	.76	-.19	.25
R - W	.50	-.03	.29
N - W	.36	-.02	.15

The correlations presented in Table 4 were computed by taking all the possible nonredundant differences between subtests at each grade level and computing product-moment  $r$ 's for corresponding pairs. The median  $r$  for the total group is .35 with a range from .14 (V minus S) to .42 (V minus R). These data are certainly not encouraging support for the use of profile differences. Further examination of Table 4 shows a higher median  $r$  for boys (.55) than for girls (.08) which was to be expected in the light of our previous findings concerning consistency of performance. A tentative explanation for these sex differences is that the academic program pursued by the boys is more in conformity with the abilities measured on the PMA at both grade levels whereas for the girls the commercial courses have the effect of changing their relative strengths and weaknesses in Grade 11 in contrast to what they were in Grade 8.

The generally lower correlation of difference scores noted for the PMA in contrast with the DAT requires further examination. The most likely explanation for these divergent results lies in the different statistical techniques used for analysis. Doppelt and Bennett (1951) computed their meas-

ures of relationship by means of a formula<sup>3</sup> which assumes equal variability among the variables, an assumption for which they do not offer supporting evidence. Judging from the data presented in the manual for the DAT, however, it would appear that the variance in Grade 12 would have been considerably larger than in Grade 9. If the assumption were seriously violated however their approach would yield an overestimate of the degree of relationship. Since our data did not fulfill the assumption of homogeneity of variance Pearson product-moment correlations were used. The median magnitude of relationship for the total group, using the Doppelt-Bennett approach, would have been .55, ranging from .46 (R minus W) to .68 (S minus N), which represents a substantial improvement over the results derived from product-moment  $r$ 's.

#### *Consistency of Factorial Structure*

In the analysis presented in this section our primary concern is with the degree to which the factor loadings on each of the PMA subtests change over time, i.e., factor stability. This problem is of importance not only for the theoretical issues involved but also in terms of more practical considerations. To our knowledge empirical data have not previously been reported on this issue for the PMA.

The scores of the 96 subjects who were present for both administrations of the PMA and the MRT were intercorrelated for boys and girls separately and for the total group. Each of the  $11 \times 11$  matrices was factor analyzed by the complete centroid method with

$$^3 r_{(1-2)(I-II)} = \frac{r_{1I} + r_{2II} - r_{1II} - r_{2I}}{\sqrt{2 - 2r_{12}} \sqrt{2 - 2r_{I\bar{I}}}}$$

where (1-2) and (I-II) represent the difference obtained at first and second testings, respectively.

TABLE 5

FIRST- AND SECOND-ORDER REFERENCE VECTOR MATRICES AND INTERCORRELATIONS  
AMONG FIRST- AND SECOND-ORDER PRIMARY MENTAL ABILITIES FACTORS

Grade Level	Scales	Total (N = 92)						Boys (N = 46)						Girls (N = 46)					
		V	S	R	N	W	$h^2$	V	S	R	N	W	$h^2$	V	S	R	N	W	$h$
8	PMA-V	62	-01	-04	-01	09	39	54	03	-05	04	14	32	69	02	-12	-01	11	50
	PMA-S	-02	67	05	05	-02	45	-05	63	10	08	-05	42	-04	72	02	03	09	53
	PMA-R	05	-01	63	-06	00	40	18	-05	53	-12	-19	37	-01	69	-01	14	50	
	PMA-N	11	-06	-03	58	-02	35	11	-10	01	60	-02	38	-13	18	02	61	09	43
	PMA-W	11	-07	04	-05	58	36	09	02	02	-04	68	47	13	-11	18	02	41	23
12	PMA-V	51	03	-04	16	02	29	62	00	-09	-01	00	39	36	19	-07	37	-09	32
	PMA-S	05	68	-06	-05	03	47	11	59	-06	-07	01	37	08	66	03	12	-01	46
	PMA-R	-07	01	64	10	04	43	-17	05	71	07	16	57	-08	07	62	16	-05	42
	PMA-N	-07	04	06	63	-01	41	-03	08	01	71	-03	51	-18	04	12	68	-07	52
	PMA-W	-05	07	01	01	61	38	-01	-02	10	02	62	40	-05	21	01	-02	46	26
Correlations among factors	HS Ach.	47	02	12	-13	-16	28	43	02	-07	02	-01	19	48	-04	26	-20	-15	36
	V																		
	S		46	58	60	38			45	68	62	28			48	49	57	10	
	R				36	10	18			31	14	10				41	24	31	
	N					56	29				57	03				46	19		15
Second-order vector loadings	A	29	51	02	-22	20		39	48	-01	-03	37		10	56	24	-03	49	
	B	13	-27	36	57	05		35	-09	61	60	-13		52	-05	32	60	-21	
Factor intercorrelation						87				66							65		

six orthogonal factors being extracted.<sup>4</sup> The fifth centroid factor showed, in each matrix, one or more loadings greater than .30, while none of the loadings on the sixth factor reached .20. Consequently, only the first five factors were retained in the subsequent rotation. The centroid factor loadings were rotated to oblique simple structure using the oblimax analytic rotation criterion developed by Pinzka and

<sup>4</sup> The authors are indebted to Gary Lotto and William B. Kehl for providing the facilities of the University of Pittsburgh Computation and Data Processing Center for part of the statistical analyses. A table giving the unrotated centroid factor loadings and transformation matrices has been deposited with the American Documentation Institute. Order Document No. 6544 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Saunders (1954) and the correlations (direction cosines) among the oblique reference vectors were converted into factor intercorrelations by the procedure detailed by Cattell (1952, pp. 224-232). Two centroid second-order factors were extracted from the factor correlation matrices and rotated to oblique simple structure using the same procedures.

Table 5 gives the oblique vector loadings of each test, the correlations among the primary factors, the vector loadings of the five primary factors on two second-order factors for each sex group and also for the combined groups. The primary factors within each analysis are clearly Thurstone's V, S, R, N, and W factors, but the factor intercorrelations are quite large in several cases, indicating that the PMA factors cannot be considered as orthogonal or independent. The presence of a single factor at the second-order level might

be considered as evidence of a general g factor unifying the five primary factors, but it appears that two factors must be postulated at the second-order level to account adequately for the factor inter-correlations at the first-order level. The large correlations between the second-order factors suggest a general g factor is present but at the third-order rather than at the second-order level.

Certain interesting sex differences in the factor structure of the PMA scales are apparent in the separate sex group analysis that are obscured when the groups are combined. The V scale appears to be less well defined at the eleventh grade level for the girls (loading of .36) although its loading is quite high at the eighth grade level (.69) and at both grade levels for the boys (.54 and .62). The V scale also has a considerable loading on the N factor (.37) at the twelfth grade level for girls although none of the other scales similarly load on two primary factors in either matrix. Second-order Factor A has loadings on both the S and W primary factors and Factor B on the V, R, and N primary factors for both sex groups. However, Factor V divides its variance between Factors A and B for the boys while V is loaded only on B for the girls. In spite of these apparent sex differences the intercorrelations between Factors A and B are quite similar in both sex groups (.66 and .65), but this correlation is much larger when all subjects are combined into the total group (.87).

Considering first the results for the total group it seems quite clear that a high degree of consistency among the PMA factors prevails from Grade 8 through Grade 11. It can be concluded from these data that the structure of intelligence changes very little over the time period included in this study. The large correlation between the factors suggests that the structure of intelligence is not composed of independent

traits but rather that these traits share a common source of variation. The foregoing interpretation is consistent with Vernon (1950) who proposes a hierarchical structure of intelligence with Spearman's (1927) g as the major integrating factor.

Considering next the data for the boys and girls, it can be concluded that factorial stability is greater for boys than for girls. That factorial structure changes for the girls as reflected in the lower factor loading on V in the eleventh grade as compared with its loading in the eighth grade is not surprising in view of the previously reported findings. A tentative explanation of these findings will be discussed in a later section of this paper.

#### *Developmental Changes in Magnitude of Sex Differences and Rates of Growth of Each of the Primary Abilities*

Herzberg and Lepkin (1954) report comparisons between boys and girls at each of three age levels (16, 17, 18) on the subtests of the PMA. At each age level the S subtest was significantly higher for the boys. Scores on the subtests V, R, and W were significantly higher for the girls at age 17 but not at age 16, and at age 18 only W was reliably higher for the girls. In evaluating these data it should be noted that all the subjects were high school seniors so that the three age levels actually represent different intellectual levels, thus confounding any conclusions concerning sex differences at various levels of development. The present data are not hampered by this limitation.

Sex differences on each of the subtests were determined for each grade level by means of *t* tests. As shown in Table 6, none of the differences were statistically significant in Grade 8 but being significantly different in favor of the girls. Though scores on the space

TABLE 6  
SEX DIFFERENCES ON PRIMARY MENTAL  
ABILITIES SUBTESTS AT GRADES  
8 AND 11

PMA Subtest	Grade 8 Means			Grade 11 Means		
	Boys	Girls	t	Boys	Girls	t
V	16.83	17.89	.75	26.65	30.63	1.99*
S	20.87	18.59	1.18	30.69	28.50	1.06
R	12.24	14.19	1.63	17.02	19.72	2.29*
N	15.19	17.73	1.63	19.61	24.28	2.52*
W	29.28	33.00	1.73	36.78	42.06	2.29*

\* p < .05.

factor favor the boys at both grades, the differences are not statistically significant. To the degree that our Grade 11 subjects are comparable to the 16-year-olds of Herzberg and Lepkin, the correspondence of findings would permit the conclusion that eleventh grade girls achieve higher scores on V, R, N, and W than boys.

#### Long Term Prediction of Achievement

This section is concerned with the effectiveness of each of the subtests to predict achievement 3.5 years later. The correlations presented in Table 7 are based on Grade 8 subtest scores and Grade 11 achievement as well as Grade 11 subtest scores and Grade 11 achievement. It will be recalled that the measure of achievement is performance on the MRT. The magnitude of these *r*'s is fairly impressive, particularly the V subtest for boys. Generally these data are in agreement with other studies (Shinn, 1956; Thurstone, 1958; Wellman, 1957), particularly with respect to the overall superiority of V in predicting achievement. The sex differences in predictive power agree with Shinn (1956) who found small but consistent differences in the magnitude of *r* in favor of boys. That our differences are larger is probably attributable to the greater proportion of our girls being enrolled in the commercial

TABLE 7  
CORRELATIONS OF PRIMARY MENTAL ABILITIES SUBTESTS ADMINISTERED AT GRADES 8 AND 11 WITH RUCH HIGH SCHOOL ACHIEVEMENT TEST ADMINISTERED AT GRADE 11

PMA Subtest	Boys (N = 46)		Girls (N = 46)		Total (N = 92)	
	Grade 8	Grade 11	Grade 8	Grade 11	Grade 8	Grade 11
V	.70	.74	.56	.45	.58	.52
S	.33	.26	.23	.29	.30	.28
R	.53	.49	.36	.34	.36	.32
N	.54	.46	.20	.12	.31	.20
W	.24	.22	.17	-.05	.14	.05

program wherein preparation for a test such as the MRT is less effective than a more academically oriented course of study.

Since it is ordinarily expected that concurrent validity is higher than long term validity, the lower correlations between eleventh grade achievement and eleventh grade PMA scores than between eighth grade PMA scores and eleventh grade achievement requires some consideration. After discarding several psychological hypotheses as being illogical or untestable, the writers next examined the data for possible statistical explanations. One possibility considered was the difference in variances from Grade 8 to Grade 11 on the PMA subtests but, it will be recalled, the Grade 11 variances are greater which should have led to larger correlations. Still attending to the difference in variances, a testable statistical hypothesis was developed. Considering the definition of *r* as the ratio of the covariance to the product of the standard deviations of the two variables<sup>5</sup> it can be shown that if the

<sup>5</sup> This formula symbolically is

$$r = \frac{\Sigma xy}{\sigma_x \sigma_y}$$

from which it can be shown that the ratio of any pair of correlations is proportional to the ratio of the standard deviations.

covariance remains constant and the  $\sigma$  of one of the two variables remains constant then any increase in the size of the second  $\sigma$  must reduce the magnitude of the resulting correlation coefficient. In the present problem the  $\sigma$  remains constant for the achievement test variable but not for the PMA subtest variable which is larger for the eleventh grade than the eighth. The crucial question remaining is whether or not the covariation between Grade 8 subtest scores and Grade 11 achievement scores is similar to the covariation between Grade 11 PMA scores and Grade 11 achievement scores. Direct computation of the essential covariances indicated that in every case they were equal or approximately equal. These latter findings then permit the conclusion that the observed differences in predictive validity are merely statistical artifacts. Nevertheless it should be noted that the greater variability in the eleventh grade PMA scores does not improve their predictive power. We believe this occurs because the harder items at the older age level are less appropriate discriminators for a test such as the MRT. Certainly further research is necessary, using a variety of achievement measures, before these assertions will have any generality.

#### DISCUSSION

Of all the conclusions warranted by this study the most surprising are those involving sex differences. The fact that scores on the subtests V, R, and W were significantly better for the girls is consistent with other data, but that these girls perform reliably better than the boys on N is clearly inconsistent with other studies (Herzberg & Lepkin, 1954). In addition the consistently higher correlations found for the boys in contrast to the girls is unique in the literature. This latter finding is prob-

ably best explained statistically in that the performance of the boys is more variable in Grade 11 than that of the girls and this, combined with the higher mean performance of the eleventh grade girls, suggests the presence of a ceiling effect for the girls. As a basis for a tentative explanation of the observed sex differences, we postulate that the training received by the girls in the commercial program leads to changes in the organization of intellectual abilities which do not result from the academic program of study followed by the boys. Recall that for the eleventh grade girls the V factor splits its variance with the N factor indicating that these factors overlap. We postulate that this overlap represents the emergence of a clerical ability which reflects an emphasis on training for accuracy and speed that is not necessarily stressed to the same degree in the academic program. But it is exactly these abilities which are important to good performance on the N subtest and to a lesser degree on the V subtest. If our analysis is correct, then the superior performance of the girls on the N subtest reflects their greater training on this type of task. This explanation is consistent with the broad cultural interpretation usually given similar data.

The writers recognize that the foregoing assertions are highly speculative but in the absence of any data of a similar nature we have attempted to put forth an idea that might be tested in a subsequent study. It may be that other interpretations will prove more fruitful but in any event with the large numbers of girls enrolled in commercial some attempts be made to replicate our findings.

Though the purpose of this study did not include an evaluation of the PMA, such a statement would appear war-

ranted. Certainly the test-retest correlations for the Factors V, R, and N are sufficiently high to permit cautious long term counseling. This is fortunate since those factors correlate most highly with achievement. The Factors S and W are quite unstable and appear to contribute little towards the prediction of future behavior. It should be noted here that despite the claims made, the V subtest is still the best predictor of achievement, is the most stable over time, and, importantly, correlates highest with the other subtests and with total score both concurrently (.73) and long term (.64). The second best predictor of achievement is the T score (.47) which is also the most stable score (.82). Though long term consistency in relative position is satisfactory, the same cannot be said for the consistency of differences between subtests. Profiles as derived from the PMA would appear to have little value in differential prediction or for counseling purposes. On the basis of the several studies reporting sex differences on the PMA subtests it seems quite clear that separate norms for boys and girls are needed if the test is to be used intelligently. Finally we should like to call for further research on the reasons for observed sex differences and for factorial validation studies employing samples other than those pursuing academic programs in the high schools and colleges.

### SUMMARY

A sample of 100 eighth graders was administered the Primary Mental Abilities Test, Intermediate Form, and was readministered the same test as they were concluding the eleventh grade. Several analyses of the resulting data were performed which showed that relative position on the PMA subtests is fairly well maintained over time but that differences between sub-

test scores are quite unstable. Though no sex differences were noted at Grade 8, there were significant differences in favor of the girls at Grade 11 on the V, R, N, and W subtests. In addition to these sex differences it was noted that girls were less consistent in performance than the boys and their scores correlated less well with achievement, as measured by the Myers-Ruch High School Achievement Test. The Verbal Meaning subtest predicted achievement most effectively, with the Total score ranking second. Analyses were also made relevant to the problem of the structure of intelligence and the degree to which specific factors emerge. The absence of evidence for increased differentiation of abilities as noted from Grade 8 to Grade 11 and the highly correlated second-order factor was viewed as support for Vernon's hierarchical-structure theory of intelligence.

### REFERENCES

- ANASTASI, ANNE. The nature of psychological 'traits.' *Psychol. Rev.*, 1948, 55, 127-138.
- ANASTASI, ANNE, & DRAKE, J. D. An empirical comparison of certain techniques for estimating the reliability of speeded tests. *Educ. psychol. Measmt.*, 1954, 14, 529-540.
- ASCH, S. A study of change in mental organization. *Arch. Psychol.*, 1936, No. 195.
- CATTELL, R. B. *Factor analysis*. New York: Harper, 1952.
- DOPPELT, J. E., & BENNETT, G. D. A longitudinal study of the differential aptitude test. *Educ. psychol. Measmt.*, 1951, 11, 228-237.
- GARRETT, H. E. A developmental theory of intelligence. *Amer. Psychologist*, 1946, 1, 372-378.
- GARRETT, H. E., BRYAN, ALICE I., & PERL, R. The age factor in mental organization. *Arch. Psychol.*, 1935, No. 176.
- HERZBERG, F., & LEPKIN, M. A study of sex differences on the Primary Mental Abilities Test. *Educ. psychol. Measmt.*, 1954, 14, 687-689.
- MEYER, W. J. The stability of patterns of primary mental abilities among junior

- high and senior high school students. *Educ. psychol. Measmt.*, 1960, 20, 795-800.
- PINZKA, C., & SAUNDERS, D. R. *Analytic rotation to simple structure: II. Extension to oblique solutions*. Princeton, N.J.: Educational Testing Service, 1954.
- SCHILLER, B. Verbal, numerical, and spatial abilities of young children. *Arch. Psychol.*, 1934, No. 161.
- SHINN, E. O. Interest and intelligence as related to achievement in tenth grade. *Calif. J. educ. Res.*, 1956, 7, 217-220.
- SPEARMAN, C. *The abilities of man*. New York: Macmillan, 1927.
- THURSTONE, L. L. Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1.
- THURSTONE, L. L., & THURSTONE, T. G. Factorial studies of intelligence. *Psychometr. Monogr.*, 1941, No. 2.
- THURSTONE, THELMA G. *Manual for the SRA Primary Mental Abilities*. Chicago: Science Research Associates, 1958.
- TYLER, LEONA. The stability of patterns of primary mental abilities among grade school children. *Educ. psychol. Measmt.*, 1958, 18, 769-774.
- VERNON, P. *The structure of human abilities*. New York: Wiley, 1950.
- WELLMAN, F. E. Differential prediction of high school achievement using single score and multiple factor tests of mental maturity. *Personnel guid. J.*, 1957, 35, 512-517.

(Received August 22, 1960)

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

April 1961

Number 2

## SEX ROLE AND COMMUNITY VARIABILITY IN TEST PERFORMANCES<sup>1</sup>

CARSON McGUIRE<sup>2</sup>

*University of Texas*

In the Human Talent Project, which has dealt so far with boys and girls in their junior high school years, one major problem has been the identification of suitable measures which combine effectively to explain and forecast talented behavior among adolescents (Hindsman & Duke, 1960). The proposed psychological model for research in human talent (McGuire, 1960) postulates three categories of variables. Each kind of behavior to be explained or predicted is, in large part, a function of (a) potentialities of an individual pertinent to that behavior, (b) expectations regarding supportive or nonsupportive responses of self and others, and (c) pressures imposed upon the person by parents, age-mates, and teachers. The pressures form a part of what Goethals (1958) terms the "context" in his framework for educational research. Not only the interpersonal but also multipersonal situations or institutional settings have to be taken into account.

<sup>1</sup>The research reported herein was supported through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare.

<sup>2</sup>The author is indebted to Earl Jennings and F. J. King (now at Florida State University), research associates of the Human Talent Project, and to Kathleen Silva, statistical clerk, Laboratory of Human Behavior, for valuable assistance in carrying out the numerous computations for this research.

This paper reports a series of tests of the proposition that sex role and school location have a moderating influence upon the performances of junior high school students on cognitive and noncognitive instruments. Each of the test variables has been selected as a possible measure of the potentialities, expectations, or pressures. If the results support the proposition, the model should be modified by two kinds of moderator variables. One is variability of certain test variables in terms of sex-role identification and the sex-typing of socialization pressures upon boys and girls (More, 1953). The other represents differences in community context and patterns of educational experiences from one location to another with accompanying variations in what is expected and what may be learned (Ferguson, 1954). Indirectly, the series of studies permits an evaluation of the usefulness of the context variables suggested by Goethals (1958). More directly, the results provide some necessary data about relations among variables which may be used to define the dimensions of and to predict talented behavior.

### METHOD

The research team was quite aware of the literature on a wide range of sex and other human differences, much of it ably summarized by Tyler (1956) and by Anastasi (1958). No studies of individual differ-

ences, however, provided data upon the complex interplay among variables encountered when subjects of both sexes were drawn from schools in several communities and when they represented different family backgrounds and levels of mental function. Therefore three different tests of the proposition were designed, each using a multivariate approach made possible by programming analyses for a modern high speed computer.

### *Analyses of Variance*

The first design made provisions for two other sources of variation in test performance in addition to sex role and school context. One of these components was family status, representing variations in culturally-typed experiences and the socialization pressures of different family backgrounds in each community (Auld, 1952; McGuire, 1953). The other was level of mental function, to test the interaction of cognitive functioning with the other three components of variance. Sex role (S), family status (F), mental function (M), and school location (L) were combined in a  $2 \times 3 \times 3 \times 4$  factorial design with two replications in each cell. The replications provided an error term with 72 degrees of freedom for testing main effects and interactions among the postulated sources of variation in performances upon 22 cognitive and 22 noncognitive test variables.

*Subjects.* The subjects were 144 junior high school students, 72 boys and 72 girls, drawn by random procedures to fit the research design from a total population of 1,417 who had responded to each of the tests in the seventh grade. Each of the four communities, designated A, B, C, and D, participating in the Human Talent Project was represented by 36 subjects. Of these boys and girls, equal numbers were selected from high (HFS), middle (MFS), and low (LFS) status homes as determined by an index of social status for each family (McGuire & White, 1955). Each subsample was subdivided for level of mental function, the measure of intelligence being the California Test of Mental Maturity (CTMM), Junior High Level, Form S, 1957. Based upon standard deviations from the mean for the total population of Anglo-, Negro-, and Latin-American students, the subdivisions were IQ 113 and above (HMF), 84 to 112 (AMF), and 83 or less (LMF).

*Cognitive variables.* Among the 22 cognitive test variables, 11 were measures of

achievement during the seventh grade. Reading, language, and arithmetic achievement were assessed by the California Achievement Tests (CAT), Junior High Level, Form W, 1957. CAT Reading had two subtests, Vocabulary and Comprehension; CAT Language had subscores for Mechanics of English and for Spelling; and CAT Arithmetic combined Fundamentals and Reasoning. The remaining achievement instruments were STEP Social Studies and STEP Science from the Sequential Tests of Educational Progress, Cooperative Test Service, Form 3A, 1957.

The other 11 cognitive test variables measured the cognitive, perceptual, and psychomotor potentialities of each boy and girl in the seventh grade year. STEP Listening, which required comprehension of passages read aloud, measured efficiency in the apprehension of verbal stimuli. Clerical Aptitude and Mechanical Reasoning were chosen from the well-known battery of Differential Aptitude Tests (DAT), Form A, 1947. Vocabulary Completion and Gestalt Transformation, the latter to estimate ability to shift the function of a part of an object and use it in a new way, were taken from a battery of factor tests used by Guilford, Wilson, Christenson, and Lewis (1951). Mutilated Words, Gestalt Completion, Short Words, and Copying were selected from a kit of reference tests, supplied by the Educational Testing Service, to evaluate various aspects of perceptual closure. Dotting and DRT, to measure psychomotor speed and discrimination reaction time, were used by special permission of the Air Research and Development Command, Lackland Air Force Base.

*Noncognitive variables.* The 22 noncognitive instruments were paper-and-pencil tests designed to assess motivations and other elements of personality as well as attitudes expressed in self reports. Scores for 11 test variables were obtained from an administration of the IPAT Junior Personality Quiz (JPQ), 1952. Of the JPQ scales, only the one for Intelligence has not been included in the analyses. Two parts of the Texas Cooperative Youth Study, developed by Moore and Holtzman (1958) for a statewide survey, provided scales to assess family tensions, negative social orientation, authoritarian discipline, personal maladjustment, criticism of education, criticism of youth, social inadequacy, and self-inadequacy. The children's form of the Manifest Anxiety Scale (CMAS), as used by Castaneda, McCandless, and Palermo (1956),

was modified for administration to adolescents (odd-even reliability .90 for 41 items). Finally, the Brown-Holtzman Survey of Study Habits and Attitudes (SSHA) was adapted for use in secondary schools (odd-even reliability .95 on 55 items keyed for boys and girls). Thus modified, the SSHA yielded two test variables, Scholastic Motivation and Teacher Valuation.

*Analyses of data.* Distributions of scores on the 44 test variables were in the form of stanine values 1 to 9 with 5 as the mean value for the total population of 1,417 students. The transformation of raw scores into stanine values not only permitted maximum use of the IBM 650 facilities at the University Computing Center but also assured additivity and homogeneity of variance. Means for the sample population hovered about a stanine value of 5 with somewhat larger deviations than the total population since the subjects represented extremes of family status and mental function as well as boys and girls in the middle range from each community.

The 44 analyses of variance were carried out on the IBM 650 computer and *F* tests were used to identify significant main effects and interactions. Comparisons were made with the deviation term (SFMLP), which had 72 degrees of freedom, unless a lower-order interaction was significant at the .01 or .05 level of confidence. Then intraclass correlation coefficients were computed for each significant mean square in the resulting tables, employing the formula.

$$r_I = \frac{MS_a - MS_w}{MS_a + (n - 1)MS_w}$$

where:

$MS_a$  = mean square for "among groups"

$MS_w$  = mean square for "within groups"

$n$  = number of subjects in each group

and using the deviation term (SFMLP) as  $MS_w$  for each computation. The intraclass correlation coefficients were regarded as measures of the degree of "resemblance" among subjects in the same sex role (S), family status (F), level of mental function (M), school location (L), or sharing some combination of these attributes.

#### Factor Analyses

The second approach to testing the proposition that sex role and school location influence performances on cognitive and noncognitive tests involved a series of

factor analytic studies to map out relationships among variables. By this time, the remaining factor tests from the Guilford battery had been scored for the total population of 1,417 males and females in the four school locations. Thus stanine values for Unusual Uses, Consequences, Common Situations, and Seeing Problems were added to the master deck along with subject grades assigned by teachers for the year and grade point averages (GPA).

In each analysis, correlations among the variables provided a matrix for the extraction of centroid factors which were rotated to an orthogonal normal varimax solution (Kaiser, 1958) by appropriate programming. The varimax criterion for analytic rotation was elected since it had the greatest likelihood of portraying factors invariant under changing samples of tests and populations. For this study, analyses were carried out for each sex role (m, f) and for each community (A, B, C, D) in addition to the one for the total population.

#### Multiple Regression Analyses

The third step involved multiple regression studies with selected criteria of academically talented behavior as dependent variables. These criterion measures were GPA Teacher Evaluation, CAT Reading, CAT Language, CAT Arithmetic, STEP Social Studies, and STEP Science. Separate analyses were carried out for each sex (m, f) in each of the four communities (A, B, C, D). The cognitive and noncognitive variables which showed promise as indicators of talented behavior were retained as dimensional or independent variables. In addition, CTMM Mental Function and ISS Family Status were added to represent differences in intellectual functioning and in family background, respectively, which were shown to be involved in the interactions found in the initial variance analysis. Rhymes, another of the Guilford et al. (1951) factor tests to measure word fluency or verbal facility, also was included. By this time, the meanings of 46 kinds of descriptions age-mates made of one another in response to nomination items had been sorted out by factor analytic and factor matching techniques (Hindsman, 1960). Thus five sociometric factor variables common to boys and girls were included in the regression studies: namely, Peer Acceptance, Absence of Negative Model Value, Social Effectiveness, Nondeviant vs. Deviant Behavior, and Quiet Dependency.

An iterative technique (Greenberger & Ward, 1956) was employed for the multiple regression analyses. The technique, which is a modification of the Kelly-Salisbury method, was adopted since the independent variables were so numerous. Using the IBM 650 computer, iteration was carried out to that point in the program where the sums of squares of regressed values,  $R^2$ , were not raised more than a specified criterion value of .0005 in the analysis. The stop criterion provided a solution which avoided overfitting the regression line. This solution was regarded as one which would have minimum shrinkage upon application of the regression weights to subsequent samples. Variables with regression weights of zero at the conclusion of the iterations were regarded as linear combinations of those with weights. Thus it was not necessary to specify in advance the subsets of independent variables which were linear combinations of the others. This was regarded as an important justification for the method employed.

## RESULTS

The outcomes of the three studies are summarized in a series of tables which clearly show variability in test performances between boys and girls and from one community to another. Many of the findings in the variance analyses confirm what is already known. What is new largely involves the significant interactions. The factor analytic and multiple regression studies demonstrate the consequences of the sex-typing of socialization pressures and the influences of school location, or community context, upon ways in which variables combine to map out different kinds of behavior and to explain certain kinds of talent valued in junior high schools.

### *Variance in Test Performances*

Table 1 shows results of the analyses of variance for performances of 144 subjects on the 22 cognitive test variables. In a similar manner, Table 2 gathers together the results for the 22 noncognitive measures. Instead of the usual mean squares, the entries

are interclass correlation coefficients which measure the resemblance or average degree of similarity among persons classified alike. For example, in Table 1, the coefficient .49 in the column for mental function (M) and row for CAT Reading Achievement has replaced a mean square significant at the .01 point in the original analysis of variance. Reference to tables of means, not shown in this report to conserve space, reveals mean stanine values of 7.9 for HMF, 5.2 for AMF, and 4.3 for LMF subsamples. The coefficient,  $r_{11} = .49$ , represents the degree to which 48 subjects in the same category of mental function are alike in performances on the standard test. Each entry in the two tables may be interpreted in a similar manner.

*Cognitive tests.* As one would expect, Table 1 shows that, without exception, mental function is a source of variation in performances on achievement, cognitive, and psychomotor tests. DAT Clerical Aptitude (speed and accuracy), Gestalt Completion (visual figural recognition), and Dotting (psychomotor speed), however, appear only in interaction with other components of variance. In general, when the coefficients of resemblance are low, as in the case of Mechanical Reasoning and Discrimination Reaction Time, the subjects in AMF and LMF subsamples are more like one another than they are similar to HMF students who exceed their performances.

Independently of mental function, family status clearly is a source of variation in some cognitive performances. The HFS boys and girls have the advantages of upper, upper-middle, and middle-middle class family backgrounds. They resemble one another more than they do either the AFS or LFS subjects in various aspects of reading, in mechanics of Eng-

TABLE 1  
SIGNIFICANT INTRACLASS CORRELATION COEFFICIENTS FROM ANALYSES OF VARIANCE IN COGNITIVE PERFORMANCES OF 144 JUNIOR HIGH STUDENTS

Cognitive Variable	Source of Variation <sup>a</sup>				
	S	F	M	L	Interaction
CAT Reading Achievement		08**	49**		
Vocabulary		05*	44**	05*	
Comprehension		08**	38**		
CAT Language Achievement	25**		41**	09**	
Mechanics of English	18**	09**	45**	13**	
Spelling	18**		29**		
CAT Arithmetic Achievement			40**	20**	
Fundamentals			29**	31**	SFML = 36*
Reasoning			39**		
STEP Social Studies		05*	29**		FML = 22*
Science			19**		FML = 29**
Listening			31**		FML = 24*
DAT Clerical Aptitude	34**				SFML = 31*
Mechanical Reasoning	19**		11**		
Vocabulary Completion	07*		49**	07*	FML = 31**
Gestalt Transformation			12**		
Mutilated Words			19**		
Gestalt Completion					SML = 19*
Short Words	16**		15**	07*	
Copying		10**	14**		SML = 21*
Dotting	15**				ML = 11*
DRT Reaction Time	06*		07*		
Subjects within group, n	72	48	48	36	
df	1	2	2	3	

Note.—Rounded and decimal points omitted.

<sup>a</sup> Sex role (S), family status (F), CTMM mental function (M), school location (L); number of subjects in each group, ML = 12, FML = 4, SML = 6, SFML = 2.

\*  $p < .05$ .

\*\*  $p < .01$ .

lish, in solving science problems, and in perceiving and copying figures accurately.

Sex role apparently has a moderating influence upon certain kinds of cognitive test performances in the first year of junior high school. Girls resemble one another and excel boys in various aspects of language achievement, in the speed and accuracy with which they cope with clerical tasks, in identifying short four-letter words embedded in rows of letters (speed of perceptual closure), and in psychomotor speed measured by a dotting test. In addition, the girls appear to have somewhat greater verbal com-

prehension, as measured by Vocabulary Completion, and Discrimination Reaction Time than the boys in the seventh grade. On the other hand, boys out perform girls in reasoning through pictorially presented mechanical situations.

Differences in test performances traceable to varying patterns of educational experiences from one community to another are most apparent in the case of achievement in arithmetic. In terms of the sample populations drawn, subjects from the two Gulf Coast communities, C and D, excel those of the North Central cities, A and B. The variability between pairs

of locations appears to be a function of mastery of arithmetic fundamentals, not ability to reason numerically. In language achievement, subjects in Community C have the highest mean scores and those in A the lowest, B and D being similar. The differences may be traced to the relative mastery of the mechanics of English language by students in the four locations. Moreover, the small but significant variations from place to place in speed of perceptual closure, as measured by Short Words, parallel the performances on language tests. On the other hand, the communities rank C, A, D, B on the two vocabulary tests variables.

Each of the significant interactions in Table 1 involves school location as one of the components of variance, and four of the nine reflect sex-typed influences. In two instances, mastery of the fundamentals of arithmetic and clerical speed and accuracy, pairs of boys and girls with similar intelligence and family backgrounds in the same community (SFML) resemble each other significantly. Performances on the three STEP tests, Social Studies, Science, and Listening (cognitive apprehension), as well as Gestalt Transformation (conceptual redefinition), vary from community to community according to family background and quality of mental function (FML) of the junior high students. Aspects of perceptual closure, measured by Gestalt Completion and the Copying test, also vary from location to location, but this time in terms of sex-typed learning experiences and intellectual functioning (SML). Finally, psychomotor speed, measured by the Dotting test, has a different relation to quality of mental function from one place to another (ML) independently of the tendency of girls to excel boys in the four communities.

*Noncognitive tests.* The analyses of noncognitive performances summarized in Table 2 refer only to self-report or paper-and-pencil tests. Responses of seventh grade students to noncognitive instruments apparently are less frequently influenced by variations in intellectual functioning than performances on cognitive tests. The 48 subjects in the average or AMF subsample resemble one another in having more nervous tension (JPQ 2) and reported anxiety symptoms (CMAS) than those of either high or low intelligence. Thus the relation of these two variables to mental function (CTMM) is curvilinear rather than linear. The relation of a preference for authoritarian discipline (CYS) to intelligence also is somewhat curvilinear, mean stanine values being 4.6 for HMF, 5.4 for AMF, and 5.2 for the LMF subsamples.

Only in two instances does family status have an independent effect upon distributions of scores obtained from the self-reports. The subjects from MFS homes, lower-middle and mobile upper-lower class, resemble one another in having less independent dominance (JPQ 9) than the 48 subjects in either the HFS or LFS subsamples. As expected, the lower class boys and girls in the LFS subsample tend to express a negative orientation to society (CYS), whereas subjects in the other two categories are alike in being positive.

A number of sex-typed differences may be noted in the noncognitive test behavior. Girls in the sample population represent themselves as being emotionally sensitive, surges or talkative and excitable, high in socialized morale or acceptance of school and dominance, tolerant and slow to anger. On the other hand, boys resemble one

TABLE 2

SIGNIFICANT INTRACLASS CORRELATION COEFFICIENTS FROM ANALYSES OF VARIANCE  
IN NONCOGNITIVE PERFORMANCES OF 144 JUNIOR HIGH STUDENTS

Noncognitive Variable	Source of Variation <sup>a</sup>				
	S	F	M	L	Interaction
JPQ 1 Sensitivity/Toughness	48**				FL = 10*
2 Tension/Relaxation			07*		SL = 18*
3 Emotional/Stable					
4 Control/Casualness					
5 Impatient Dominance	06*				
6 Sociable/Withdrawn					
7 Adventurous					
8 Socialized Morale	20**				
9 Independent Dominance	13**	05*		10**	
10 Energetic Conformity					
11 Surgency/Desurgency	24**				SFML = 62**
CYS Family Tension					
Negative Social Orientation		10**			
Authoritarian Discipline			06*		SFML = 40*
Personal Maladjustment					SFM = 18*
Criticism of Education					
Criticism of Youth					SL = 23**
Social Inadequacy					FML = 32**
Self Inadequacy					SFM = 22*
					FML = 30**
CMAS Anxiety			05*		
SSHA Scholastic Motivation					SFML = 41*
Teacher Valuation	05*				FL = 09*
Subjects within group, n	72	48	48	36	
df	1	2	2	3	

Note.—Rounded and decimal points omitted.

<sup>a</sup> Sex role (S), ISS family status (F), CTMM mental function (M), school location (L); number of subjects in each group, SL = 16, FL = 12, SFM = 8, FML = 4, SFML = 2.

\*  $p < .05$ .

\*\*  $p < .01$ .

another more than girls in being tough-minded, more serious or desurgent, reacting negatively to learning tasks and school authority, independent, impatient, and not so appreciative of their teachers. Sex role, family status, and mental function (SFM) combine to influence responses to the CYS scales for criticism of education and feelings of self-inadequacy. In terms of the self-reports, the eight lower class girls of high intelligence, two from each community, are least critical and feel least inadequate. Most critical of the school and its expecta-

tions are boys of low intelligence from lower-middle and mobile upper-lower class families. Feelings of self-inadequacy are highest among lower class boys who also are low in mental function.

School location is an independent source of variation in only one set of responses to the noncognitive instruments: namely, independent dominance which is highest among the 36 subjects in Community C and lowest in D. On the other hand, 9 of the 11 significant interactions in Table 2 involve variability from one community

to another. The most interesting ones are representations of social inadequacy (SL, FML), self-inadequacy (FML), and family tension (SFML), all departing from zero at the .01 level of confidence. For example, the 18 boys in Community B express the greatest social inadequacy (SL). The least social and self-inadequacy on the CYS scales is represented by the four lower class subjects of high intellectual calibre in Community A (FML). Pairs of boys or girls with similar level of mental function and family status in the same community (SFML) have a remarkable resemblance in CYS scores for family tension ( $r_I = .62$ ).

None of the cognitive test variables and only six of the noncognitive instruments remain uninfluenced by one or more of the components of variance included in the research design. They are JPQ 3, 4, 6, 7, and 10, together with CYS values for criticism of youth. Of these, only JPQ 3 (emotionality vs. stability) and 6 (sociable vs. withdrawn) later proved to be promising indicators of talented behavior (McGuire, Hindsman, Jennings, & King, 1961). As a main effect or in an interaction, there are sex-typed differences in performances on 12 of the 22 cognitive tests and on 13 of the 22 noncognitive measures. Similarly, there are either independent or complex variations in responses from one community to another for 15 of the cognitive and 9 of the noncognitive variables.

#### *Factor Analytic Studies*

Table 3 presents a sample of pertinent loadings for cognitive and noncognitive variables from the factor analytic studies employed to sort out abilities and other attributes of the seventh grade population and a num-

ber of subpopulations.<sup>3</sup> The table is arranged so that magnitudes of loadings for the two sex roles (m, f) and among the four communities (A, B, C, D) may be compared for Factor B (academic grades) and Factor D (socially oriented achievement motivation). The number of subjects for each analysis is entered at the base of each column in the table. Following convention, loadings of .30 or greater are entered for the variables defining each factor. The letter a denotes the highest loading of a variable for the population studied.

*Sex roles.* As shown in Table A (deposited with ADI), the paper-and-pencil test responses and teacher evaluations of the 1,417 junior high school students, 772 males and 645 females, can be mapped out into seven factors. Factor matching of loading magnitudes (Cattell, 1957, pp. 818-827) indicates a reasonably close agreement (correlations of .70 or greater over all variables) in the case of five of the seven factors for boys and girls. The factors common to both sexes have been provisionally named Scholastic Achievement (largely defined by standard tests of achievement), Academic Grades (reflecting the teacher evaluations of performances in the seventh grade), Divergent Thinking (three of the four proposed tests of creativity), Ineffectively Functioning Personality (self-reports of maladjustment, inadequacy, tension, and anxiety), and Social

<sup>3</sup>A four-page table (Table A) giving comparisons of factor loadings from 50-variable matrices for total population, sex role, and school location has been deposited with the American Documentation Institute, Order Document No. 6610 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies.

TABLE 3  
FACTOR LOADINGS OF COGNITIVE AND NONCONGITIVE VARIABLES FOR GRADE  
VII POPULATION, BY SEX ROLES, AND COMMUNITIES

Variable	Grade	Sex Roles		Communities			
		VII	m	f	A	B	C
Factor B: Academic Grades							
GPA Teacher Evaluation		82 <sup>a</sup>	82 <sup>a</sup>	81 <sup>a</sup>	86 <sup>a</sup>	—	79 <sup>a</sup>
Teacher Social Studies		71 <sup>a</sup>	70 <sup>a</sup>	70 <sup>a</sup>	80 <sup>a</sup>	—	65 <sup>a</sup>
Teacher English		70 <sup>a</sup>	69 <sup>a</sup>	69 <sup>a</sup>	84 <sup>a</sup>	—	63 <sup>a</sup>
Teacher Mathematics		65 <sup>a</sup>	69 <sup>a</sup>	63 <sup>a</sup>	68 <sup>a</sup>	—	75 <sup>a</sup>
CAT Arithmetic		—	—	30	33	—	32
CAT Language		—	—	—	36	—	44
STEP Social Studies		—	—	—	—	—	42
CAT Reading		—	—	—	—	—	36
STEP Reading		—	—	—	—	—	30
DAT Science		—	—	—	—	—	32 <sup>a</sup>
JPQ 3 Emotional/Stable		—	—	—	-31	—	—
Factor D: Socially Oriented Achievement Motivation							
JPQ 8 Socialized Morale		71 <sup>a</sup>	72 <sup>a</sup>	75 <sup>a</sup>	75 <sup>a</sup>	47	66 <sup>a</sup>
JPQ 9 Independent Dominance		-57 <sup>a</sup>	-52 <sup>a</sup>	-52 <sup>a</sup>	-52 <sup>a</sup>	-44 <sup>a</sup>	-57 <sup>a</sup>
JPQ 3 Emotional/Stable		-53 <sup>a</sup>	-48 <sup>a</sup>	-52 <sup>a</sup>	-35 <sup>a</sup>	-60 <sup>a</sup>	-47 <sup>a</sup>
SSHA Scholastic Motivation		47 <sup>a</sup>	48 <sup>a</sup>	49 <sup>a</sup>	38 <sup>a</sup>	39 <sup>a</sup>	38 <sup>a</sup>
JPQ 6 Sociable/Withdrawn		42 <sup>a</sup>	40 <sup>a</sup>	39 <sup>a</sup>	40 <sup>a</sup>	38 <sup>a</sup>	42 <sup>a</sup>
CYS Critic of Education		-38	—	40	-35	-32	-30
CMAS Anxiety		-32	—	-39	—	-63 <sup>a</sup>	—
JPQ 1 Sensitive/Tough		30 <sup>a</sup>	—	—	50 <sup>a</sup>	—	32
JPQ 11 Surgent/Desurgent		—	-36 <sup>a</sup>	-36	—	—	-32
CYS Negative Social Orientation		—	—	—	—	-44	—
GPA Teacher Evaluation		—	—	—	—	31	—
Teacher Social Studies		—	—	—	—	30	—
CYS Social Inadequacy		—	—	—	—	-63 <sup>a</sup>	—
CYS Maladjustment		—	—	—	—	-74 <sup>a</sup>	—
CYS Family Tension		—	—	—	—	-66 <sup>a</sup>	—
Population, N		1,417	772	645	338	319	448
							264

Note.—Decimal points omitted.

<sup>a</sup> Highest loading for a variable in a factor structure.

Alienation (authoritarian and anti-social attitudes).

The two factors with least agreement between the sex roles turn out to be Socially Oriented Achievement Motivation (acceptance of school and cultural standards, conformity, stability, and scholastic motivation) and Perceptual Closure. Factor loadings for the former, Factor D, are shown in

Table 3. Girls who have a positive attitude toward school and academic attainments are less anxious than boys and appear to be able to be critical of educational practices whereas their male counterparts are not. Elements common to both sexes on the perceptual factor involve ability to cognize symbolic units (Mutilated Words, Short Words, DAT Clerical). For girls,

this ability is linked with achievement on a language test; for boys, symbolic closure is tied in with visual figure recognition and psychomotor speed.

*Community variability.* With one exception, Social Alienation, the factors mapped out for the total population fail to match across the four communities. The factor structure for Community C, which had 488 boys and girls in departmentalized classrooms of two junior high schools, is most like that of the total population. In the other school locations, Perceptual Closure is absorbed either by Divergent Thinking (A and D) or by Scholastic Achievement (B). Community A, which had 338 students in modified self-contained classrooms of one large junior high, otherwise shows a pattern which is reasonably similar to C and the total population. In Community D, where 264 pupils were housed in self-contained classrooms of county schools as well as in a smaller departmentalized junior high school, Academic Grades appear to be more closely associated with performances on standard tests of achievement than in Locations A and C.

The factor analysis for Community B, where seventh grade students were housed in self-contained classrooms at five different elementary schools pending the completion of another junior high building, yields a sorting of abilities and other attributes quite different from Locations A, C, and D. Only four factors are necessary to map the paper-and-pencil test responses and teacher evaluations of the 319 boys and girls. High loadings for the standard tests of achievement and for teacher evaluations combine with two measures of divergent thinking (*Seeing Problems, Consequences*) and indicators of perceptual closure to form a composite factor which might be called Academic Attainment. Thus for Factor B (Academic Grades), there

are no loadings shown for Community B in Table 3. Factor D (Socially Oriented Achievement Motivation) in Table 3 for Community B becomes what might be termed "effectively vs. ineffectively functioning personality," a polar arrangement of JPQ, CYS, and similar noncognitive scales. Only Social Alienation remains similar to the other communities.

### Regression Studies

Table 4, which indicates the regressions of CAT Language Achievement in Grade 7 upon certain cognitive and noncognitive variables by sex role in each of the four communities, serves to illustrate the results of the regression studies. Only the beta weights for variables which illustrate between sex and cross community variability most effectively are shown. Similar variability of beta weights for variables between the sexes from community to community occurs when the criterion measures are CAT Reading, CAT Arithmetic, STEP Social Studies, STEP Science, and GPA Teacher Evaluation.<sup>4</sup> When GPA Teacher Evaluation is the dependent variable, the sociometric factor variables and some of the noncognitive indicators of talented behavior have higher loadings than they do for standard tests of achievement used as criteria. Although the loading magnitudes for the set of variables vary with each cri-

<sup>4</sup> A set of six tables (Tables B to G) showing regressions of CAT Reading, CAT Language, CAT Arithmetic, STEP Social Studies, STEP Science, and GPA Teacher Evaluation by community and sex role has been deposited with the American Documentation Institute. Order Document No. 6610 from ADI Auxiliary Publications Project, Photoduplication Service, Library in advance \$1.25 for microfilm, remitting for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 4

REGRESSIONS OF CAT LANGUAGE ACHIEVEMENT IN GRADE VII UPON COGNITIVE  
AND NONCOGNITIVE VARIABLES BY COMMUNITY AND SEX ROLE

Variable <sup>a</sup>	Community							
	A		B		C		D	
	m	f	m	f	m	f	m	f
Multiple correlation, <i>R</i>	.81	.82	.82	.84	.81	.83	.75	.81
Sample population, <i>N</i>	161	157	94	112	266	241	113	98
CTMM Mental Function	39	30	28	17	33	20	41	29
DAT Mechanical	.07	.04	20	-.06		.08	.19	-.12
Gestalt Transformation			-.08		.21	.06	.06	.08
Common Situations	-.05	-.06	.05	-.08	-.12	-.09	.11	-.12
Short Words	.15	.11	.25	.03	.06	.10	.08	.11
Dotting	.04	.11	-.07	.13	.06		-.16	.08
JPQ 1 Sensitivity	-.07	.05	-.06	-.03		.12	.10	
JPQ 8 Social Morale		-.13		.03	-.03	-.08	.11	
JPQ 11 Surgency			.04	.12	-.03		-.07	.16
SSHA Scholastic Motiv'n	.05	.09	.19		.05	.11	-.17	
N-2 Model Value	.12	.14	-.10					.18
ISS Family Status		.05		.20			-.14	

Note.—Rounded to two places, decimal points omitted in lower portion of table.

<sup>a</sup> Beta weights for all 32 variables are shown in Table C, deposited with the American Documentation Institute.

tion, the multiple correlation coefficients are uniformly high and range from .75 to .86 (McGuire et al., 1961).

The important point to be noted in Table 4, and in the complete Tables B to G (deposited with ADI), is that the subset of independent variables which yields the maximum multiple correlation with the criterion measure is not the same from one community to another, nor for males and females at each of the school locations. In each column of the complete tables, the omitted variables (which have beta weights of zero) can be regarded as linear combinations of those with loading magnitudes for the particular subpopulation. Thus the regression studies supply further evidence that the sex-typing of socialization pressures and the context of educational experiences apparently do have quite an influence upon ways in which cognitive and noncognitive variables rep-

resenting potentialities, expectations, and pressures combine to explain various kinds of valued behavior.

#### DISCUSSION

The series of variance, factor analytic, and regression studies of sex role and community variability in test performances, as well as teacher evaluations and peer appraisals, confirm some well-known findings and raise questions about others. For example, the variance analyses of seventh grade data fail to support the generally held conclusion that females excel in numerical computation and males are superior in mathematical reasoning and in science (Anastasi, 1958, pp. 452-504; Tyler, 1956, pp. 247-275). Although the reputed aggressiveness of boys can be read into their greater tough-mindedness, independence, and negative reactions to learning tasks and school authority,

the seventh grade males are more serious (desurgent) than their female counterparts in this study. The higher emotional sensitivity and excitability (surgecy) of the girls might be interpreted as symptoms of neuroticism and instability, but this is belied by their tolerance, socialized morale, and restraint.

Taken together, the series of studies clearly support the proposition that sex role and school location have a moderating influence upon the performances of junior high school students on cognitive and noncognitive instruments. In the variance analyses, more than half of the responses appear to be biased by whatever is involved in learning a male or a female sex role. The consequences are more evident in the regression than the factor analytic studies, where five of seven factors showed a reasonable match between the sexes. Within communities, however, the differences between males and females are quite marked where the same criterion measure regresses on different sets of independent variables with fluctuating beta weights.

Similarly, more than half of the distributions for cognitive tests, and nearly half for the noncognitive instruments, vary from one school location to another. These variations are illustrated most clearly by the differences among communities in the mastery of fundamentals of arithmetic and mechanics of English. More important, however, are the frequent interactions of school location with other components of variance. They would lead one to infer not only that patterns of educational experiences differ but also that more subtle influences are being exerted by the four community contexts. These inferences are borne out both by the differences in factor structure, which parallel de-

gree of departmentalization or organization in terms of self-contained classrooms, and by the regression studies.

One of the unexpected findings is the frequency with which life style is independent of intelligence in influencing behavior on cognitive tests, especially standard measures of achievement. The most probable explanation of the additive relation between social class and mental function is that high family status is an index of the kinds of motivation and socialization pressures which maximize experiences involving reading, the effective use of English, and concern about scientific knowledge. This interpretation parallels a finding in another study (McBee & Duke, 1960) that intelligence and SSHA scholastic motivation have independent effects upon achievement in arithmetic, reading, and science, but not language and social studies.

The foregoing findings posed a very difficult problem for the research team. They suggested that the sex-typing of socialization pressures and the influences of school location upon ways in which the proposed classes of variables (potentialities, expectations, pressures) combined were more important than first postulated. If so, a unique combination of indicators to explain and predict talented behavior would have to be worked out for males and females of each community. The alternative was to estimate the underlying factors in persons (Cattell, 1957, pp. 287-296; Guilford, 1954, pp. 524-526) to attain what Guilford terms "a much more standard article."

#### SUMMARY

A series of variance, factor analytic, and regression studies were undertaken to test the proposition that sex role and school location have a moderating influence upon the perform-

ances of junior high school students on cognitive and noncognitive instruments selected as potential indicators of talented behavior. Using a factorial design which involved 144 boys and girls from four communities in their seventh grade year, analyses of variance were carried out for 22 cognitive and 22 noncognitive variables. These analyses demonstrated many expected and some unanticipated main effects of differences in sex role, family status, mental function, and school location. More important, they showed 20 significant interactions, 18 of which involved variations in test behavior from one community to another. Then the consequences of sex role and community variability in test performances, first identified in the variance analyses, were demonstrated in factor analytic studies and multiple regression analyses.

The regression analyses revealed that various kinds of behavior valued in junior high schools could be explained with multiple correlations ranging from .75 to .86 by varying combinations of cognitive and noncognitive variables. But these sets of independent variables to be used as predictors were different for each community and for boys and girls within each school location. This development paralleled the finding that factor structures varied for males and females, and even more so from one community to another.

## REFERENCES

- ANASTASI, ANNE. *Differential psychology*. (3rd ed.) New York: Macmillan, 1958.
- ✓AULD, F. Influence of social class on personality test responses. *Psychol. Bull.*, 1952, **49**, 318-332.
- CASTENADA, A., McCANDLESS, B. R., & PALERMO, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, **27**, 317-326.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- FERGUSON, G. A. On learning and human ability. *Canad. J. Psychol.*, 1954, **8**, 95-112.
- GOETHALS, G. W. A framework for educational research. *Harv. Educ. Rev.*, 1958, **28**, 29-43.
- GREENBERGER, M. H., & WARD, J. H., JR. An iterative technique for multiple correlation analysis. *IBM Tech. Newslet.*, 1956, No. 12, 85-97.
- GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.
- ✓GUILFORD, J. P., WILSON, R. C., CHRISTENSON, P. R., & LEWIS, D. J. A factor-analytic study of creative thinking: I. Hypotheses and description of tests. *Rep. Psychol. Lab., U. Southern Calif.*, 1951, No. 4.
- HINDSMAN, E. Dimensions of adolescent behavior. Unpublished doctoral dissertation, University of Texas, 1960.
- HINDSMAN, E., & DUKE, R. L. Development and utilization of talent. *J. teacher Educ.*, 1960, **9**, 107-116.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
- ✓MCBEE, G., & DUKE, R. L. Relationships between intelligence, scholastic motivation, and academic achievement. *Psychol. Rep.*, 1960, **6**, 3-8.
- ✓McGUIRE, C. Family and age-mates in personality formation. *Marriage fam. Liv.*, 1953, **15**, 17-23.
- ✓McGUIRE, C. *Talented behavior in junior high schools*. Final Report, 1960, University of Texas, Project 025, Cooperative Research Program, Office of Education, United States Department of Health, Education, and Welfare.
- ✓McGUIRE, C., HINDSMAN, E., JENNINGS, E., & KING, F. J. Dimensions of talented behavior. *Educ. psychol. Measmt.*, 1961, **21**, 3-38.
- McGUIRE, C., & WHITE, G. D. The measurement of social status. *Res. Pap. hum. Develpm.*, 1955, No. 3.
- MOORE, BERNICE M., & HOLTZMAN, W. H. What Texas knows about youth. *Nat. Parent-Teacher*, 1958, **53**, 22-24.
- ✓MORE, D. M. Developmental concordance and discordance during puberty and early adolescence. *Monogr. Soc. Res. Child Develpm.*, 1953, **18**(1, Ser. No. 56).
- TYLER, LEONA A. *The psychology of human differences*. (2nd ed.) New York: Appleton-Century-Crofts, 1956.

# SIGNIFICANT IQ CHANGES IN TWENTY-FIVE YEARS:

## A FOLLOW-UP<sup>1</sup>

KATHERINE P. BRADWAY  
*Stanford University*

AND

NANCY M. ROBINSON<sup>2</sup>  
*Chapel Hill, North Carolina*

The following constitutes a report of several statistical analyses of intelligence test and family background data obtained in 1931, 1941, and 1956, on 110 members of the original pre-school standardization sample of the 1937 Revision of the Stanford-Binet.<sup>3</sup> An initial report of the study has appeared previously (Bradway, Thompson, & Cravens, 1958).

In 1931, 212 children in the San Francisco Bay Area, aged 2 to 5½, were given tests which later constituted Forms L and M of the 1937 Stanford-Binet. Careful methods of selection were adhered to in selecting these subjects, as they were the California sample of the nationwide standardization of this revision (McNemar, 1942; Terman & Merrill, 1937). Ten years later, the senior author administered Form L of the Stanford-Binet to the 138 of these children who still remained in the area.

In 1956, 110 of those tested in 1931 and 1941 were located and given both Form L of the 1937 Stanford-Binet and the Wechsler Adult Intelligence Scale. No interim contact had been

<sup>1</sup>This investigation was supported by PHS Research Grant M-1273 (C2) from the National Institutes of Health, United States Public Health Service.

<sup>2</sup>Formerly at Stanford University.

<sup>3</sup>The authors are grateful to Maud A. Merrill and the late Lewis M. Terman for generously having made their standardization data available for follow-up; to Quinn McNemar, consultant for both follow-ups, and to Clare Thompson for her several contributions to the handling of data and to their presentation.

made between 1931 and 1941, or between 1941 and 1956.

An important aspect of the 1941 study involved those 50 subjects whose test scores showed statistically significant increase or decrease since the original (1931) testing. Extensive interviews with their parents were conducted to assess the intellectual stimulation in the home and to obtain data on the intellectual level of the parents and grandparents. Significant differences between the groups were shown on an "ancestral index" and related measures which estimated the intellectual abilities of parents and grandparents (Bradway, 1945). Included in the 1956 testing were 21 of the 26 subjects whose IQs had shown the most increase between 1931 and 1941 ("1941 increase group"), and 21 of the 24 subjects whose IQs had decreased most during that period ("1941 decrease group"). It is these 42 subjects who figure most prominently in the analyses to follow, which were designed to determine whether a change between preschool and junior high school age is more likely to be followed in adulthood by a change in the same direction, the opposite direction, or no change at all in the individual's status compared to the rest of the group. A further question involves the previous finding that a significant decrease in IQ was associated with a significantly lower intellectual level in the two previous generations than was a significant increase in IQ.

### SUBJECTS AND PROCEDURE

The subjects in the present study had been administered Forms L and M of the

Revised Stanford-Binet Scale 25 years previously, in 1931, in connection with its standardization, and Form L of the same scale in 1941 (Bradway, 1945). A comparison of the 1941 retest group of 138 subjects with the total 633 subjects of like age in the original United States standardization group showed some upward selection. The initial mean composite IQ of the retest group was 109.2, compared with 105.4 for the standardization group.

In selecting for further study the 54 subjects whose IQs had changed most, special corrections were applied to the initial composite IQs. These allowed for regression of the retest (1941) IQ due to errors of measurement of the initial (1931) IQ, and equated initial and retest IQs in relation to the means of their distributions. The details of these adjustments have been reported by Bradway (1945).

In obtaining information in 1941 about home environment, not only were the 50 children interviewed, but also mothers of 47 and fathers of 3 of them. The mothers, in addition, were given a vocabulary test of intelligence and each child was given the Woodworth-Cady Questionnaire. Twenty-six scores were obtained for a wide variety of variables such as intellectual stimulation of the home, intelligence of parents, grandfathers' intelligence as estimated by the Minnesota Scale of Occupational Intelligence (Brussell, 1932), happiness and socialization of the household, and child's general adjustment. The most consistent and marked differences occurred in scores reflecting the intelligence of parents and grandparents. An index which gave weight both to parental intelligence and grandfathers' occupational intelligence produced a critical ratio of 3.32 ( $p < .001$ ) between the groups, the 26 children in the "increase" group coming from families with significantly higher index values than the 24 children in the "decrease" group.

The loss of 28 subjects between 1941 and 1956 testings resulted in further upward selection with respect to initial (1931) mean IQ, which was 111.1 for the 1956 group, or 5.7 points higher than the mean for the total standardization group of these ages. At the time of the 1956 testing, subjects were interviewed about their occupation, marital status, education, and health. Information about parents' education and grandfathers' occupation was gathered by mail and is available for 98 of the 110 subjects. Ancestral indices were calculated using methods

similar to those employed by Bradway in the previous study.

## RESULTS

### *Test Scores for 1941 Groups*

An earlier article (Bradway et al., 1958) reported a correlation ( $r$ ) of .59 between preschool Stanford-Binet IQs and adult Stanford-Binet IQs obtained 25 years later; the correlation between adolescent and adult IQs obtained in 1941 and 1956 for the same group was .85. They also found that the obtained 1956 Stanford-Binet IQs were higher than would be expected on the basis of the 1931 and 1941 IQs of this group, although the standard deviations of the distributions were similar.

Presumably, intellectual growth does not cease by the age of 16 as had been assumed in the 1937 Revision (Terman & Merrill, 1937). This group had attained a mean mental age of 18.5 years as adults, instead of 16.8 which would have been predicted from their adolescent IQs in the absence of mental growth beyond age 16 years. This finding is in correspondence with the reports of several recent investigators that intellectual growth continues well past adolescence (Bayley, 1955; Bayley & Oden, 1955; Miner, 1957; Owens, 1953). To equate for this continued growth and to permit meaningful comparisons between an individual's status within the group at the two testings, the mean difference between the 1941 and 1956 IQs (11.5 points) was subtracted from the obtained 1956 Stanford-Binet IQs.<sup>4</sup> The logic underlying this correction is similar to that upon which the use of the Deviation IQ is based: namely, that the IQ is intended as an index of the individual's relative standing within his age group, rather than as a measure

<sup>4</sup> A forthcoming article by Bradway and Thompson will consider adult intelligence and the problems of appropriate indices.

of his level of mental development per se. There was no reason to suspect that these particular subjects had exhibited unusual mental growth which would have elevated the mean IQ of this sample above the mean IQ of the general population; and their Deviation IQs on the WAIS tended to substantiate this conclusion. Using standard scores would have been an alternate way of equating the two sets of scores, but since the standard deviations of the two sets of IQs were similar, equating the means by subtracting the mean difference from each of the 1956 IQs was sufficient to make them comparable to the 1941 IQs as indices of relative position within the group.

The means and standard deviations of the resulting corrected IQs of the present sample are included in Table 1, arranged according to the subgroups to which they were assigned in 1941. In this table are presented also means of the Stanford-Binets administered in 1931 and 1941 and of the WAIS administered in 1956 to the same groups. The respective mean chronological ages at the times of these

three testings were 4.0, 13.6, and 29.5 years.

It will be noted from Table 1 that for both 1941 change groups, the observed changes to 1956 are in the reverse direction from the previous changes. It would seem that the group whose rate of mental growth had shown deceleration from preschool to adolescence, however, had more nearly reached a level of brightness by adolescence that was largely maintained into adulthood. Subtracting their 1941 IQs from their 1956 (corrected) IQs yielded a *t* value for this decrease group of 1.94, with a probability between .05 and .10. On the other hand, the group whose mental growth had shown acceleration between preschool and adolescence had tended to lose some of their relative gains by the time they reached adulthood. The corresponding *t* value for this increase group was -3.02, which is significant at beyond the .01 level. Moreover, the difference between the 1956 mean IQs for the two groups is reliably less than the difference between the 1941 mean IQs for the two groups as indi-

TABLE 1  
MEAN IQ AND STANDARD DEVIATION OF 1941 SUBGROUPS FOR EACH TEST

Test	1941 Subgroups			
	Decrease (N = 21)	Increase (N = 21)	No change (N = 68)	Total (N = 110)
1931 Forms L, M				
Mean IQ <sup>a</sup>	112.8	108.8	111.3	111.1
SD	21.0	12.8	15.1	15.9
1941 Form L				
Mean IQ	97.3	127.9	111.7	112.0
SD	17.2	11.0	13.7	16.8
1956 Form L				
Mean IQ (Corr.) <sup>b</sup>	100.6	121.5	112.5	112.0
SD	18.6	9.5	13.1	15.1
1956 WAIS				
Mean IQ	101.8	113.5	108.8	108.4
SD	14.0	7.6	9.1	10.6

<sup>a</sup> Average of two IQs minus one point to correct for practise effect.

<sup>b</sup> IQ minus 11.5 points to equate 1941 and 1956 group means.

cated by a  $t$  value of 3.59 which is significant at beyond the .001 level.

An index of the stability of the scores is provided by determining whether, in fact, the groups remain significantly different from each other, as they were in 1941. Analyses of variance of the 1956 data yielded an  $F$  of 12.28 for the Stanford-Binet IQs and an  $F$  of 7.40 for the WAIS Full Scale IQs, with probabilities of <.001 and .001, respectively. The groups do, then, remain significantly different from each other.

#### *Further Investigation of the "Ancestral Index"*

In an effort to establish the generality of the finding that ancestral intelligence is related to the direction of change in mental growth, each subject was asked by mail to indicate how far each of his parents had gone in school, and the occupation of each of his grandfathers. This information was already available for those subjects of the 1941 "change" groups, but was not known for the 68 subjects who had been members of the "no change" group. The mean initial (1931) IQ of the 98 subjects for whom these data became available was 112.3. The mean educational level of their parents was 11 years, which is comparable to the mean educational level of California residents according to the United States Bureau of the Census (1950).

In the 1941 study, the weighted index had been calculated on ranked data for the 50 subjects in the "change" groups, giving equal weight to the following: (a) mothers' intelligence as ranked by interviewer (largely based on Vocabulary test score), (b) fathers' intelligence as ranked by interviewer (largely based on occupation), and (c) average of grandfathers' occupational scores on the Minnesota occupational classifica-

tion (Brussell, 1932). Ranks on each variable were translated to scores on a normalized distribution according to a method suggested by Hull (Garrett, 1932, p. 113). Because the data consist of ranks within this selected sample only, the absolute scores cannot be translated directly to other groups.

A similar method was used in calculating the ancestral index for the 1956 data. In this case, however, the rank order of mothers' intelligence was estimated solely on the basis of education, an index admittedly inferior to Vocabulary test score, but the best index available to us. A rank order correlation of .72 was found between education and interviewer estimates (largely based on Vocabulary score) for the 40 subjects in the deviation groups for whom both figures were obtained in 1941. Fathers' intelligence was estimated by education and Minnesota occupational classification, equally weighted. Grandfathers' occupations were classified as before. Each rank was translated into a score according to the Hull method, and the scores were combined to give equal weight to mothers', fathers', and grandfathers' (combined) scores.

For the 28 subjects in the "change" groups who returned the questionnaire, a correlation of .70 was obtained between ancestral indices based entirely on the 1956 information, and those based on 1941 information. Since it seemed probable that the accounts by the mothers were more accurate than those given 17 years later by the younger generation, the 1941 information was utilized wherever possible.

The resulting indices were then correlated with each set of test data, as shown in Table 2. Whatever the nature of the relationship between ancestral index and intelligence, it appears that the main effects have occurred by the time of adolescence. The cor-

TABLE 2  
CORRELATION OF IQS WITH  
ANCESTRAL INDEX  
(*N* = 98)

Test	<i>r</i>	<i>r/sigma</i> <sup>a</sup>
1931 Combined Forms L, M	.14	1.38
1941 Form L	.29	2.86**
1956 Form L	.23	2.27*
1956 WAIS-Full Scale	.32	3.15**
1956 WAIS-Verbal	.31	3.05**
1956 WAIS-Performance	.29	2.86**

<sup>a</sup> Standard error of the correlation is .1015 when *N* is 98.

\* Significant at .05 level.

\*\* Significant at .01 level.

relation of the ancestral index with IQ did not quite reach statistical significance ( $p = >.1$ ) in 1931 (preschool level); it was significant in 1941 (adolescent level); it remained significant, and without appreciable change, in 1956 (adult level). The *t* tests of the differences between these related correlations were calculated according to a method suggested by McNemar (1955, p. 146). The difference between correlations of the ancestral index and test scores obtained in 1931 and 1941 reached a borderline level of significance ( $p = .07$ ), but none of the correlations obtained in 1941 and 1956 differed significantly from each other.

### DISCUSSION

Our data indicate that mental growth continues into adulthood (as others have reported). More specifically, however, we found that a deceleration of mental growth rate from preschool to adolescence is not likely to be followed by further (permanent) changes in growth rate up to adulthood. On the other hand, acceleration of mental growth rate from preschool to adolescence is more likely to be followed by deceleration (of a much smaller degree than the previous ac-

celeration) than by further acceleration or by no change at all.

The findings that the factors related to the rate of mental growth had their greatest effect during the earlier years of life is consistent with findings of the several other longitudinal studies which have correlated various parental and environmental factors with variation in children's IQ. For example, Bayley (1940) in the Berkeley Growth Study and Honzik (1940) in the Berkeley Guidance Study reported that the correlation between various economic and cultural indices in the home increased from 18 months to, but not beyond, various early childhood ages.

This study was not, of course, approached with the expectation of shedding light on the fast-extinguishing nature-nurture controversy. The hope was, however, to establish some practical rules-of-thumb to aid the diagnostician who is called upon to make an estimate of a child's future mental status. While no precise recipe has been derived, these data seem to indicate that the psychologist who tests a preschool child would be wise to shade his estimate of future status somewhat in the direction of the general family level of intellectual ability. On the other hand, this would not hold for the junior high school counselor. So far, we have little which will improve his estimate beyond the IQ obtained under favorable testing conditions. The counselor's chances of being reasonably accurate even without such further aids, however, are considerable; as noted earlier, the correlation between the results of the administrations of Form L in 1941 and 1956 is .85 for our group.

### SUMMARY

In 1956, Form L of the Stanford-Binet, and the WAIS, were adminis-

tered to 110 young adults who had previously been given the Stanford-Binet at preschool and adolescent ages. In 1941, those subjects whose IQs had increased or decreased significantly since 1931 were selected for special attention.

Present findings:

1. For each group, regression toward the mean occurred between 1941 and 1956, but the change in this period was small compared with the previous 10 years.

2. Subjects whose IQs had increased between preschool and adolescence showed more reversal in their positions in the distribution of the total sample than did those whose IQs had decreased.

3. Significant changes in position in the distribution of IQs at preschool and the distribution of IQs of the same group at adolescence are related to ancestral intelligence; this does not hold true between adolescence and adulthood.

#### REFERENCES

- BAYLEY, NANCY. Factors influencing the growth of intelligence in young children. *Yearb. Nat. Soc. Stud. Educ.*, 1940, **39**, Part II, 49-79.
- BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805-818.
- BAYLEY, NANCY, & ODEN, MELITA H. The

maintenance of intellectual ability in gifted adults. *J. Gerontol.*, 1955, **10**, 91-107.

BRADWAY, KATHERINE P. An experimental study of factors associated with Stanford-Binet IQ changes from the preschool to the junior high school. *J. genet. Psychol.*, 1945, **66**, 107-128.

BRADWAY, KATHERINE P., THOMPSON, CLARE W., & CRAVENS, R. B. Preschool IQs after twenty-five years. *J. educ. Psychol.*, 1958, **49**, 278-281.

BRUSSELL, E. S. A revision of the Barr-Tassig Scale of Occupational Intelligence. Unpublished master's thesis, University of Minnesota, 1932.

GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1932.

HONZIK, MARJORIE P. Age changes in the relationship between certain environmental variables and children's intelligence. *Yearb. Nat. Soc. Stud. Educ.*, 1940, **39**, Part II, 185-205.

MCNEMAR, Q. *The revision of the Stanford-Binet scale*. Boston: Houghton Mifflin, 1942.

MCNEMAR, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.

MINER, J. B. *Intelligence in the United States*. New York: Springer, 1957.

OWENS, W. A. Age and mental abilities: A longitudinal study. *Genet. psychol. Monogr.*, 1953, **48**, 3-54.

TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.

UNITED STATES BUREAU OF THE CENSUS. *United States census of population: California*. 1950, P-B 5.

(Received July 14, 1960)

# A NOTE ON SCREENING SCHOOL BEGINNERS WITH THE BENDER GESTALT TEST

ELIZABETH M. KOPPITZ

*Stanwood, Mt. Kisco, New York*

VERDENA MARDIS AND THOMAS STEPHENS<sup>1</sup>

*Montgomery County Public Schools, Dayton, Ohio*

There is a growing interest in the screening of school beginners both as a means of determining the readiness for learning and as a way to spot children with potential difficulties early so that they can be given the required attention and help when it will be most beneficial. The Lee-Clark Reading Readiness Test (Lee & Clark, 1951) and the Metropolitan Readiness Test (Hildreth & Griffith, 1946) are two well standardized and widely used screening tests. The present study explores the usefulness of the Bender Gestalt test (Bender, 1938) as a screening tool for beginning first grade students.

## METHOD

**Subjects.** The subjects were 272 beginning first grade students from 11 classes in seven different schools, none were repeating the first grade. The schools selected represent a socioeconomic cross section and are located in rural, semirural, suburban, and urban areas.

**Procedure.** During the first 6 weeks of the school year eight of the teachers administered the Lee-Clark Reading Readiness Test to their classes, while three other teachers gave the Metropolitan Readiness Test to their groups. During this same period each subject was seen individually by a psychologist who administered the Bender Gestalt test. At the end of the school year the Metropolitan Achievement Test, Primary I Battery, Form R, was administered to all 11

classes by their respective teachers. The Readiness tests and the Achievement test were scored by the teachers following standard procedure for each test. The Bender protocols were scored by the psychologists according to a system developed by the senior investigator for use with school children (Koppitz, 1958, 1960).

Actual achievement as measured on the Total Average Achievement score of the Metropolitan Achievement Test was compared for each subject with his predicted achievement on each of the screening tests. Predictive scores were derived from the grade equivalent scores on the Readiness tests and from standard deviation scores on the Bender. Pearson product-moment correlations were computed between the Readiness tests and the Bender, and between the three screening tests and the Metropolitan Achievement Test.

## RESULTS AND DISCUSSION

Table 1 shows the correlations between the various tests, all of quite similar magnitude and all statistically significant. Thus it appears that the Bender can predict Total Average Achievement as measured on the Metropolitan Achievement Test as well as the Lee-Clark and the Metropolitan Readiness Test. It is of course, realized that the Lee-Clark Reading Test is specifically designed to test reading and is not meant to predict total average achievement. However, in discussing the validity of the Reading Readiness test, Lee and Clark quote Henig (1949) who compared the Reading Readiness test scores with actual reading grades at the end of the school year and obtained a correlation

<sup>1</sup>The authors wish to express their appreciation for the assistance provided by Sam Bonham and the staff of the Pupil Personnel Department of Montgomery County, as well as by the teachers and principals of the schools included in this study.

TABLE 1  
CORRELATIONS BETWEEN THE BENDER,  
READINESS TESTS, AND ACHIEVEMENT  
TEST

School	N	Lee-Clark & Met. Ach.	Bender <sup>a</sup> & Met. Ach.	Lee-Clark & Bender <sup>a</sup>
A	53	.67**	-.67**	-.64**
B	56	.42**	-.37**	-.30*
C	40	.54**	-.41**	-.21
D	26	.40*	-.58**	-.33
E	24	.67**	-.61**	-.54**
Total	199	.66**	-.68**	-.61**
		Met. Read. & Met. Ach.	Bender <sup>a</sup> & Met. Ach.	Met. Read. & Bender <sup>a</sup>
F	31	.63**	-.71**	-.73**
G	42	.66**	-.29	-.41**
Total	73	.59**	-.58**	-.59**

<sup>a</sup> All correlations with the Bender are negative since the Bender is scored for errors.

\* Significant at .05 level.

\*\* Significant at .01 level.

of .59 which is similar to the results obtained in this study.

The discrepancies between the various predictions of achievement and the actual achievement of each subject revealed some differences between the three screening tests. It appears that the Lee-Clark test tends to overestimate achievement more often than the Bender and the Metropolitan Readiness Test. The latter two tests tend to underestimate achievement more often. A follow-up investigation on those subjects who had marked discrepancies between their actual and predicted achievement suggests that the Lee-Clark test is more strongly influenced by cultural and social factors,

while the Bender reveals apparently more the potential ability in visual-motor perception which may or may not be fully developed and put to use. Further exploration of these tentative findings seems indicated.

### SUMMARY

The Bender Gestalt test and the Lee-Clark Reading Readiness Test or the Metropolitan Readiness Test, respectively, were administered to 272 beginning first grade students. Test scores were correlated with each other and with actual achievement at the end of the school year. It was found that the Bender correlates well with the Readiness tests and can predict actual achievement as well as they can.

### REFERENCES

- BENDER, LAURETTA. *A Visual Motor Gestalt Test and its clinical use*. New York: American Orthopsychiatric Association, 1938.
- HENIG, M. S. Predictive value of a reading readiness test and teachers' forecast. *Elem. sch. J.*, 1949, **50**, 41-46.
- HILDRETH, GERTRUDE. *Metropolitan Achievement Test, Primary I Battery, Form R*. Yonkers-on-Hudson: World Book, 1946.
- HILDRETH, GERTRUDE, & GRIFFITH, NELLIE L. *Metropolitan Readiness Test*. Yonkers-on-Hudson: World Book, 1946.
- KOPPITZ, ELIZABETH M. The Bender Gestalt test and learning disturbances in young children. *J. clin. Psychol.*, 1958, **14**, 292-295.
- KOPPITZ, ELIZABETH M. The Bender Gestalt test for children: A normative study. *J. clin. Psychol.*, 1960, **16**, 432-435.
- LEE, J. M., & CLARK, W. W. *Lee-Clark Reading Readiness Test*. Los Angeles: California Test Bureau, 1951.

(Received July 14, 1960)

## SOME RELATIONSHIPS BETWEEN PUPIL BEHAVIOR AND CERTAIN TEACHER CHARACTERISTICS<sup>1</sup>

DAVID G. RYANS

*University of Texas*

It usually is assumed that what goes on among pupils in a classroom is dependent to a major degree upon the presence of a teacher—a teacher whose role is to provide leadership in “learning” activities at which the pupils themselves are initiates, unprepared without assistance to determine appropriate goals or to select and employ suitable ways and means of attaining such goals. The teacher, by virtue of education and experience, is expected to be able to guide and assist his pupils in the acquisition of usable knowledge, understandings, skills, and group approved attitudes and value systems. It is unthinkable, except in very rare cases, that individual children or groups of children *entirely on their own* might successfully pursue the kinds of objectives that permit adjustment to the complex social and material worlds in which we live.

Qualitatively, at least, the logic of this argument appears sound: i.e., pupil behavior is a function of teacher behavior, the teacher being a necessary, though not sufficient, condition for purposeful and productive pupil performance. Beyond this point we usually are less willing to express ourselves regarding the nature of the pupil behavior-teacher behavior relationship. We like to think specified, distinguishable teacher acts or teacher behaviors produce, or at least contribute to, certain distinguishable pu-

pil acts or behaviors (Mitzel, 1960; Ryans, 1957, 1960). Some of us go on to postulate quantitative or functional relationships between kinds of teacher behavior and pupil behavior—assuming that more of a certain kind of teacher behavior will result in more effective pupil behavior—and proceed to gather data and evaluate the evidence (Anderson & Brewer, 1946; Christensen, 1960; Cogan, 1958; Gnagney, 1960; Jayne, 1945; Withall, 1952; and others). However, such functional relationships are not easy to demonstrate empirically, partly because of difficulties of observation and mensuration and, undoubtedly, partly because of the complex interactions among teacher and pupil characteristics (Ryans, 1956). Certainly research that might permit inferences regarding dependency relationships (producer-product relationships), with classroom situations providing the setting, is difficult to arrange and carry through—largely due to limitations imposed by practicality.

A correlational approach (less satisfactory than a dependency relations strategy for yielding the definitive kinds of answers a researcher often seeks, but nevertheless providing important information about the existence and degree of functional relationships) often poses much less of a data obtaining problem and, at the same time, provides an alternative avenue to clues and cues—in this case, regarding the influence of teacher behavior on pupil behavior.

It was this second sort of relationship (the interdependency or correlational relationship) between “what

<sup>1</sup>This research was made possible by a subvention from the Grant Foundation and was sponsored by the American Council on Education. It was reported at the 1960 meetings of the American Educational Research Association.

teachers do" and "what pupils of the teacher do" with which the presently reported investigation was concerned: with the degree to which certain indexes of pupil behavior tend to be functionally related to, or to manifest variance in common with, certain indexes of teacher behavior.

### METHOD

First of all, it is necessary to point out that the approach to pupil behavior employed in this research was a direct one, involving the immediate observation and assessment of the behavior of pupils. The pupils were observed in their regular classrooms in the presence of their teachers—teachers who *also* were observed and assessed.

Admittedly, this directly observed pupil behavior *in process* may be quite different from the products resulting from pupil behavior (Mitzel, 1960; Mitzel & Gross, 1956; Ryans, 1957, 1960). Researchers in the area sometimes indicate it is more important and relevant to judge pupil behavior in terms of its products—the measurable skills, understandings, and attitudes acquired by pupils—than merely to assess samples of ongoing pupil behavior. After all, they point out, it is the end product we are teaching for.

Now the method employed in the present research does involve an apparently reasonable, though certainly unproved, assumption: namely, that pupils who carry on their classroom activities in certain ways will acquire intended learnings more successfully than pupils who behave or act in other ways. But the fact remains that education is interested ultimately in pupil behavior and in teacher behavior for the results produced; immediate behavior is of interest only insofar as it is a means to an end. So, one might contend that the pupil behavior with

which this paper deals is of only incidental concern. The author does not agree, but it is a point of methodology that should be made clear at the outset.

Next, it is appropriate to explain what *kinds* of directly observable pupil behavior the research dealt with.

Obviously many specific behaviors, or acts, may be noted in a classroom even during a very limited sample of time. One thing assumed in this research (a basic postulate of trait theory that makes the researcher's task tolerable) was that individual behaviors possess some generality and therefore may be classified in a fairly limited number of classes or groups, each group being made up of related specific behaviors that have a common substratum which identifies the group or class. Thus, a number of pupil behaviors such as "rudeness to teacher and/or other pupils," "interruption of one another," "impatience," "refusal to participate," "quarrelsomeness," "sullenness," "disturbing noisiness," etc. might all be considered as belonging to a class of behaviors we could, for convenience, label obstructiveness. Again, pupil behaviors such as "courteousness," "friendliness," attentiveness to advice or criticism," "diligence in completing study assignments," "initiation of constructive activity in the absence of specific direction from the teacher," etc. could be thought of as belonging to a class of behaviors we might call responsibility.

In this research we selected (after considerable study of previous classroom and "personality" research and after a good deal of trial and error involving classroom observation and assessment) and focused our attention on a limited number of "important" pupil behavior dimensions. The term "dimension" was used with intent, since in planning the research it seemed reasonable to attempt to identify and assess acts of pupils in the classroom, first, in light of

their qualitative characteristics (the general class of behaviors to which a specific act belonged) and then, according to their positions on quantitative scales defined by subclasses of behaviors falling at opposite poles of the assumed continua. Thus, one bipolar dimension of pupil behavior selected because it was judged to be an important one in the classroom was the Apathetic-Alert dimension. Specific pupil behaviors that seemed to contribute to alertness (e.g., "responded eagerly," "appeared anxious to recite and participate," "watched teacher attentively," etc.) defined one end of the continuum and pupil acts that appeared to contribute to apathy ("listless," "restless," "wandering attention," "indifference," etc.) defined the other.

In attempting to assess pupil behavior,

trained observers employed several such dimensions, assigning each class of pupils observed a value, on a scale extending from 1 to 7, on each dimension. An assessment of 1 (at the left pole) on the Apathetic-Alert dimension indicated an inference presumably based upon observation of a preponderance of apathetic-type behaviors among the pupils; a 7 (at the right pole) indicated an inference based upon the presumed observation of many alert-type pupil behaviors. A "Glossary," which accompanied the assessment form and which was employed in the training of observers, helped to standardize the assessment procedure.

Estimates of the reliability of assessment of the pupil behavior dimensions, based on correlations between independent assessments of two observers visiting the same

TABLE 1  
LOADINGS OF OBSERVED DIMENSIONS OF PUPIL BEHAVIOR AND TEACHER ON A  
PUPIL BEHAVIOR FACTOR: SECONDARY SCHOOL CLASSES  
(From Ryans and Wandt, 1952)

Classroom Behavior Dimension	Oblique Factor					
	A	B	C	D	E	F
<b>Pupil Behavior:</b>						
Apathetic-Alert	.07	.01	.53	.21	.05	-.05
Obstructive-Responsible	-.02	.03	.46	.04	-.06	.10
Uncertain-Confident	.17	.02	.43	-.01	.17	-.03
Dependent-Initiating	.00	-.05	.55	.04	.21	-.02
<b>Teacher Behavior:<sup>a</sup></b>						
Partial-Fair						
Autoocratic-Democratic			-.04			
Aloof (G)-Responsive			-.02			
Aloof (I)-Responsive			.25			
Restricted-Understanding			.02			
Harsh-Kindly			.08			
Dull-Stimulating			-.02			
Stereotyped-Original			.32			
Apathetic-Alert			.17			
Unimpressive-Attractive			.12			
Inarticulate-Articulate			.01			
Monotonous (V)-Pleasant			-.04			
Evading-Responsible			-.01			
Erratic-Steady			-.08			
Excitable-Poised			.08			
Uncertain-Confident			.06			
Disorganized-Systematic			.20			
Inflexible-Adaptable			.02			
Pessimistic-Optimistic			.00			
Immature-Integrated			.19			
Narrow-Broad			.20			
			-.06			

Note.—N = 249 senior high school mathematics, science, English, and social studies classes.

<sup>a</sup> Loadings of teacher behavior dimensions are omitted here for factors on which the loadings of pupil behavior dimensions were not pronounced.

classes at different times ranged from .43 to .65 for 1,518 elementary school classes observed, and from .43 to .63 for 1,911 secondary school classes.

Teacher behavior, observed and assessed following the same procedure as that described for pupil behavior, was considered in light of dimensions such as Harsh-Kindly, Disorganized-Systematic, Dull-Stimulating, etc. (see Tables 1 and 2 below). Reliability estimates of the assessments of the several dimensions of observed teacher behavior, based on correlations of the assessments of the first observer and second observer, generally were between .50 and .60—for elementary teachers coefficients for the different dimensions clustered around .55, and for secondary school teachers, around .60.

A minimum of two independent sets of assessments was obtained for each class of

pupils and for each teacher, the separate ratings subsequently being weighted equally (by transformation to a scale employing a common mean and standard deviation) and combined to form composite assessments for each teacher and each class of pupils on each of the bipolar dimensions considered.

The teachers and classes of pupils upon which data of this report are based, generally, were from the school systems of fairly large cities. Geographically, the midwest and west accounted for a majority of the school systems sampled (93%). The classes represented 274 elementary schools and 103 secondary schools. Because of dependence upon the voluntary cooperation both of administrative offices responsible for school systems and of individual teachers, the sample was not selected so as to constitute a known probability sample.

TABLE 2

LOADINGS OF OBSERVED DIMENSIONS OF PUPIL BEHAVIOR AND TEACHER BEHAVIOR ON  
TWO FACTORS INVOLVING, IN PART, "PUPIL PARTICIPATION" AND "CONTROLLED  
PUPIL ACTIVITY": ELEMENTARY SCHOOL CLASSES  
(From Ryans, 1952)

Classroom Behavior Dimension	Oblique Factor				
	A	B	C	D	E
Pupil Behavior:					
Disinterested-Alert	.30	.50	.11	.08	.00
Obstructive-Constructive	.16	.63	-.07	.07	.04
Restrained-Participating	.40	.10	.04	.00	.18
Rude-Self-Controlled	-.01	.66	.10	.16	-.05
Apathetic-Initiating	.60	.29	.05	-.08	.02
Dependent-Responsible	.46	.49	.19	-.03	-.10
Teacher Behavior: <sup>a</sup>					
Partial-Fair	.11	.06			
Autoocratic-Democratic	.42	-.11			
Aloof (G)-Responsive	.00	-.09			
Restricted-Understanding	.27	-.06			
Unattractive-Attractive	.00	.02			
Disorganized-Systematic	-.01	.56			
Inarticulate-Fluent	.00	.33			
Inflexible-Adaptable	.36	.02			
Harsh-Kindly	.25	.08			
Apathetic-Alert	.05	.08			
Aloof (I)-Responsive	-.02	-.10			
Stereotyped-Original	.52	.14			
Changeable-Constant	-.04	.54			
Excitable-Calm	-.00	.36			
Uncertain-Confident	-.02	.44			
Irresponsible-Responsible	.04	.60			
Pessimistic-Optimistic	.05	.04			
Infantile-Mature	.21	.44			

Note.—N = 275 third and fourth grade elementary school classes.

<sup>a</sup> Loadings of teacher behavior dimensions are omitted here for factors on which the loadings of pupil behavior dimensions were not pronounced.

## RESULTS

Estimates of pupil behavior and of teacher behavior, obtained by the direct observation procedures described above, were subjected to correlational analyses. It should be noted that the findings, therefore, have to do simply with concomitance and permit inferences of interdependency relationships only. Inferences regarding "antecedent-consequent" or "producer-product" relationships may be suggested by such data, but the design employed precludes the testing of dependency relationship hypotheses.

### *Relationships among the Pupil and Teacher Behavior Dimensions*

One set of findings to be reported has to do with the extent to which the observed pupil behavior dimensions were related to the concurrently observed teacher behavior dimensions.

Here, in fairness to inferences that may be drawn, it should be noted that the pupil behaviors of the class of a given teacher were observed and assessed during the same session, or period, during which the teacher of that class was observed and assessed. Obviously there well may be contamination of one set of assessments by the other. (It would have been possible to at least partially correct for this, but it also would have added materially to both dollar and time expenditures; consequently, the choice was to recognize this source of error variance, but not to attempt to control it.)

In reviewing the pupil behavior-teacher behavior relationship findings the results for elementary teachers and secondary teachers will be presented separately; inferences that may be drawn are somewhat different. The interaction of grade level with relationships of this sort appears to be pronounced.

### *Relationships in the secondary*

*school.* In the secondary school (Ryans & Wandt, 1952), all of the pupil behavior dimensions considered in this research tended to be interrelated—the intercorrelations ranging from .42 to .68. They tended to hang together as a single cluster, all contributing to the same oblique factor when the intercorrelations of pupil behavior and teacher dimensions were factor analyzed. Table 1 shows relevant factor loadings. It may be noted from Table 1 that the factor defined by the distinctly homogeneous cluster of pupil behavior dimensions was *not* significantly contributed to by the teacher behavior dimensions assessed, with one exception (the Dull-Stimulating teacher behavior dimension). Of all the teacher behavior dimensions, *only* that which had to do with the extent to which a teacher was judged Dull-Stimulating seemed to be closely associated with the pupil behavior dimension cluster.

In the secondary school, pupil behavior (as here defined and assessed) seemed not to be as closely related to teacher behavior as we sometimes assume it to be.

*Relationships in the elementary school.* In the elementary school (Ryans, 1952) we note a distinctly different situation. Table 2 presents results comparable to those shown for the secondary school in the preceding table. With the elementary school data, however, the tight clustering of pupil behavior dimensions, relatively independent of teacher behavior dimensions (noted in the high school), is absent. Instead, the pupil behavior correlations investigated—with inter-split up and fell into one or the other or groups, of combined pupil behavior and teacher behavior dimensions.

In elementary school classrooms, the

loadings of the pupil behavior dimensions thus fell on two somewhat different factors, rather than on a single one, as was the case when secondary school classrooms were considered.

The pupil behavior dimensions contributing to the first factor appeared to be ones having to do with pupil participation. And related to this set of pupil dimensions were those teacher behavior dimensions that had been labeled Stereotyped-Original, Autocratic-Democratic and Inflexible-Adaptable—traits not entirely unlike the Dull-Stimulating teacher behavior dimension in the secondary school analysis. These kinds of relationships appear to make sense—common, as well as statistical.

The pupil behavior dimensions, which as a group might be referred to as *controlled* pupil activity, were loaded on a second factor, which also included such teacher behavior dimensions as Irresponsible (evading)-Responsible, Disorganized-Systematic, Changeable-Constant, Infantile-Mature, and Uncertain-Confident—teacher characteristics that suggest disorganized, irresponsible vs. businesslike, responsible teacher behavior.

In the elementary school, then, pupil behavior and teacher behavior seemed to be more noticeably *interdependent* than in the secondary school, with *participating* pupil behavior seeming to be related to flexible, original, democratic teacher behavior and *controlled* pupil behavior to systematic, responsible teacher behavior.

It also may be noted parenthetically that the analyses undertaken in connection with this research yielded evidence suggestive of a moderate positive relationship between the factors contributed to by the pupil behavior dimensions and those teacher behavior dimensions that loaded on a factor which seems to refer to "the way the

teacher appears" to his or her class (teacher behavior dimensions such as Monotonous-Pleasant (voice), Unimpressive-Attractive, Inarticulate-Articulate, and Uncertain-Confident contributed to this teacher appearance pattern). This sort of relationship between pupil behavior and teacher appearance seemed to hold to some degree among both elementary and secondary school classes. In the secondary school, the median correlation between the teacher behavior dimensions that contributed to teacher appearance and the pupil behavior dimensions was .27; and the correlation between the teacher appearance factor and the factor dominated by pupil behavior (Factor C of Table 1) was .33. In the elementary school, the median correlation between assessments on the teacher behavior dimensions that related to teacher appearance and assessments of the pupil behavior dimensions was .37; and the correlations between the teacher appearance factor and the factors contributed to by the pupil behavior dimensions (Factors A and B of Table 2) were .57 and .23, respectively.

#### *Relationships Based on an Overall Index of Pupil Behavior*

Now, to turn to another related aspect of the analyses which supports the foregoing findings, although coming at the problem in a slightly different manner:

In an extension of the studies just discussed, but with new independent samples of teachers and classes, an index (based on the combination of assessments for teacher behavior dimensions that loaded most highly on factors generated by the earlier factor analyses—Ryans, 1952, and Ryans & Wandt, 1952)<sup>2</sup> was obtained for each

<sup>2</sup> Assessments of each observer on each teacher behavior dimension were trans-

TABLE 3

## CORRELATIONS BETWEEN THE PUPIL BEHAVIOR INDEX AND TEACHER BEHAVIOR PATTERNS

Sample	$r_{P_0 X_0}$	$r_{P_0 Y_0}$	$r_{P_0 Z_0}$
Elementary school classes:			
834 Grades 1-6 classes	.82	.80	.75
144 Grades 1-6 classes	.83	.78	.80
Secondary school classes:			
497 Mathematics and science classes	.20	.18	.21
568 English and social studies classes	.18	.21	.26
114 Mathematics, science, English, and social studies classes	.07	.11	.14

Note.— $P_0$  = pupil behavior (composite).

$X_0$  = kindly, understanding vs. aloof, restricted teacher behavior.

$Y_0$  = responsible, businesslike vs. evading, unplanned teacher behavior.

$Z_0$  = stimulating, imaginative vs. dull, routine teacher behavior.

teacher on each of three prominent patterns of teacher behavior, and the teacher behavior pattern scores then considered in relation to pupil behavior indexes (a composite score for each class— $P_0$ —obtained from the assessments assigned the pupil behavior dimensions)<sup>2</sup> of the teachers' classes.

The teacher behavior patterns considered were ones suggested by the earlier factor analyses: (a) aloof, restricted, egocentric vs. kindly, understanding, warm teacher behavior; (b) evading, unplanned, slipshod vs. responsible, systematic, businesslike teacher behavior; and (c) dull, round-formed into standard scores. Then for each pattern of teacher behavior ( $X_0$ ,  $Y_0$ ,  $Z_0$ ) the standard scores of the several components were summed to provide the three pattern scores—based on the observer's assessments of a particular teacher. These pattern scores were, in turn, transformed into standard scores for each observer; and standard scores for independent observers were combined to yield a composite assessment on each of the behavior patterns for each teacher.

The pupil behavior index was obtained simply by (a) summing for each class, each observer's assessments (previously transformed to standard scores) on the several pupil behavior dimensions and (b) computing the mean of the equally weighted indexes contributed by the several observers who assessed a particular class of pupils.

tine vs. stimulating, imaginative teacher behavior.

Reliabilities of the indexes (scores) for the patterns of teacher behavior just named (based on correlations between the indexes yielded by assessments of different observers who independently assessed the same teachers) ranged from .70 to .80. Reliability estimates of the pupil behavior index were slightly lower—e.g., for one sample of 98 classes, when the pupil behavior dimensions were separately summed for two observers and the two sets of resulting  $P_0$  values correlated, reliability coefficients of .62, .84, .72, and .64 were obtained, respectively, when the observations were made approximately 30 minutes apart, more than 30 minutes apart but during the same half-day, during different halves of the day (forenoon or afternoon) but the same day, and at least one month apart but during the same school year.

Correlations between pupil behavior and patterns of teacher behavior based on approximately 1,000 elementary school classes and a like number of secondary school classes are presented in Table 3.

*Secondary school relationships.*  
When the teacher behavior pattern

scores were correlated with the pupil behavior index for secondary school classes of mathematics, science, English, and social studies, the evidence suggested low positive relationships between pupil behavior and teacher behavior, the obtained correlation coefficients between the pupil behavior index and the three teacher behavior pattern scores being between .07 and .26. It probably may be inferred that the relationship is not pronounced in secondary school classes.

*Elementary school relationships.* With elementary school classes, the pupil behavior index was found to be uniformly highly correlated with all three teacher behavior pattern scores (again, as suggested by the analyses reported in the preceding sections of this paper), the correlation coefficients ranging from .75 to .83. (It may be observed that these pupil behavior-teacher behavior correlations in the elementary school were as high as the reliability coefficients for the several pattern scores.)

Thus, for the kinds of teacher behavior and pupil behavior described, the two separate analyses support the same conclusion: i.e., that teacher behavior and pupil behavior show substantially more *interdependence* in the elementary school as compared with the secondary. There also is the suggestion that of the teacher behavior dimensions and patterns studied, Dull-Stimulating teacher behavior may be more closely associated with secondary school pupil behavior.

Such findings, it should be recalled, are based upon correlational analysis, permitting interdependency, but not dependency (producer-product), inferences about the relationships involved. Furthermore, the relationships reported refer to group data—to groups of teachers and their classes—and the probability risks involved in

applying conclusions to specific situations should be kept in mind.

### SUMMARY

Relationships between trained observers' assessments of (a) classes of pupils and (b) the teachers of those classes, relative to selected dimensions of pupil and teacher classroom behavior, were studied.

For elementary school classes, *high* positive relationships were noted between observers' assessments of "productive pupil behavior" (e.g., assessments presumed to reflect pupil alertness, participation, confidence, responsibility and self-control, initiating behavior, etc.) and observers' assessments of previously identified patterns of teacher behavior which seemed to refer to understanding, friendly classroom behavior; organized, businesslike classroom behavior; and stimulating, original classroom behavior.

For secondary school classes, *low* positive relationships appeared to obtain between productive pupil behavior and the above named categories of teacher behavior, with a tendency for the stimulating, original teacher classroom behavior pattern to show a slightly higher correlation with pupil behavior than the understanding, friendly or the organized, businesslike teacher behavior patterns.

The approach was correlational, permitting inferences of interdependency, but not producer-product, relationships.

### REFERENCES

- ANDERSON, H. H., & BREWER, J. E. Studies of teachers' classroom personalities: II. Effects of teachers' dominative and integrative contacts on children's classroom behavior. *Appl. psychol. Monogr.*, 1946, No. 8.
- CHRISTENSEN, C. M. Relationships between pupil achievement, affect-need, teacher

- warmth, and teacher permissiveness. *J. educ. Psychol.*, 1960, **51**, 169-174.
- COGAN, L. Behaviors of teachers and the productive behavior of their pupils. *J. exp. Educ.*, 1958, **27**, 89-124.
- GNAGNEY, W. J. Effects on classmates of a deviant student's power and response to a teacher-exerted control technique. *J. educ. Psychol.*, 1960, **51**, 1-8.
- JAYNE, D. D. A study of the relationship between teaching procedures and educational outcomes. *J. exp. Educ.*, 1945, **14**, 101-134.
- MITZEL, H. E. Criteria of teacher effectiveness. In, *Encyclopedia of educational research*. New York: Macmillan, 1960. Pp. 1481-1486.
- MITZEL, H. E., & GROSS, CECILY F. A critical review of the development of pupil growth criteria in studies of teacher effectiveness. (Research Series 31) New York: Board of Higher Education, Office of Research and Evaluation, 1956.
- RYANS, D. G. A study of criterion data: A factor analysis of teacher behaviors in the elementary school. *Educ. psychol. Measmt.*, 1952, **12**, 333-344.
- RYANS, D. G. Theory development and the study of teacher behavior. *J. educ. Psychol.*, 1956, **47**, 462-475.
- RYANS, D. G. Notes on the criterion problem in research, with special reference to the study of teacher characteristics. *J. genet. Psychol.*, 1957, **91**, 33-61.
- RYANS, D. G. Teacher characteristics in relation to marital status of the teacher. *Calif. J. educ. Res.*, 1959, **10**, 68-72.
- RYANS, D. G. Prediction of teacher effectiveness. In, *Encyclopedia of educational research*. New York: Macmillan, 1960. Pp. 1486-1491.
- RYANS, D. G., & WANDT, E. A factor analysis of observed teacher behavior in the secondary school: A study of criterion data. *Educ. psychol. Measmt.*, 1952, **12**, 574-586.
- WITHALL, J. Assessment of the social-emotional climates experienced by a group of seventh graders as they moved from class to class. *Educ. psychol. Measmt.*, 1952, **12**, 440-451.

(Received July 28, 1960)

# INVENTORY ESTIMATED TEACHER CHARACTERISTICS AS COVARIANTS OF OBSERVER ASSESSED PUPIL BEHAVIOR<sup>1</sup>

DAVID G. RYANS

*University of Texas*

In an earlier report, Ryans (1961) presented classroom observation data indicative of: (a) among elementary school classes, *high* positive relationships between observers' assessments of "productive pupil behavior" (e.g., assessments presumed to reflect pupil alertness, participation, confidence, responsibility and self-control, initiating behavior, etc.) and observers' assessments of previously identified patterns of teacher behavior which seemed to refer to understanding, friendly classroom behavior; organized, businesslike classroom behavior; and stimulating, original classroom behavior; and (b) among secondary school classes, *low* positive relationships between productive pupil behavior and the above named categories of teacher behavior, with a tendency for the stimulating, original teacher classroom behavior pattern to show a slightly higher correlation with pupil behavior than the understanding, friendly or organized, businesslike teacher behavior patterns.

If, as the evidence suggests, interdependency relationships between overt teacher behavior and overt pupil behavior in the classroom are demonstrable, a question which next is suggested is whether or not similar relationships exist between personal characteristics of the teacher as revealed by self-report inventory scores and overt pupil behavior in teachers'

classes—whether or not inventory estimated teacher traits correlate with, and therefore may be used to predict, pupil classroom behavior. It was to this problem the present research was addressed.

Two hypotheses derived from earlier studies (Ryans, 1961) were considered:

1. Certain teaching oriented, trait (or "characteristic") scores yielded by teachers' responses to a self-report inventory, the Teacher Characteristics Schedule, covary in a nonrandom manner with indexes of observer assessed, overt pupil behavior in the teachers' classes, permitting better-than-chance prediction of one from the other.

2. Teacher characteristic-pupil behavior relationships are more notable in elementary than in secondary school classes.

## METHOD

### *Pupil Behavior Data*

The classroom activities of pupils were directly observed by trained and experienced observers. The direct observation approach involved a reasonable, though certainly unproved, assumption: namely, that pupils who carry on their classroom activities in certain ways will acquire the intended "learning" (i.e., knowledge, understandings, attitudes) more successfully than pupils who behave in less seemingly productive manners.

It is important to note the particular kinds of pupil behavior that were observed and assessed—thus providing an operational definition of pupil

<sup>1</sup>This research was made possible by a subvention from the Grant Foundation and was sponsored by the American Council on Education. It was reported at the 1960 meetings of the American Psychological Association.

classroom behavior as considered in this research. Obviously, the number of specific pupil behaviors occurring in a classroom is almost unlimited. It was necessary, therefore, to make the usual assumption that such specific behaviors possess some generality and that they may be classified in a fairly limited number of groups or classes, each group being made up of related behaviors possessing a common substratum which identifies the behavior group. Thus, specific pupil acts involving "rudeness to teacher and/or other pupils," "interruption of one another," "refusal to participate," "quarrelsomeness," "disturbing noisiness," and the like might reasonably be thought of as belonging to a class of behaviors labeled, for convenience, obstructiveness. Or, specific pupil behaviors involving "courteousness," "friendliness," "attentiveness to advice or criticism," "diligence in completing study assignments," "initiation of constructive activity on one's own," and the like, might be grouped together in a class of behaviors we might call responsibility.

The groups or categories of pupil behavior selected for consideration were determined in light of review and analysis of previous classroom and personality research and after extensive pilot classroom observation and assessment during preliminary phases of the study.

For assessment purposes, the major pupil behavior categories to be considered were cast in the form of dimensions, thus, permitting the identification and assessment of observed pupil behavior (*a*) in the light of its essential qualitative characteristic (the general class of behaviors to which a specific act belongs) and (*b*) according to its position on a quantitative scale defined by subclasses of behavior falling at opposite poles of the assumed continua. Thus, one bipolar dimension of pupil behavior that observers were trained to assess was an Apathetic-Alert dimension. Pupil acts or behaviors that seemed to contribute to alertness (e.g., "responded eagerly," "appeared anxious to recite and participate," etc.) defined one pole

of the dimension and behaviors contributing to apathy (e.g., "listlessness," "restlessness," "wandering attention," etc.) defined the other.

Four such dimensions of pupil behavior were observed and assessed: Apathetic-Alert, Obstructive-Responsive, Uncertain-Confident, and Dependent-Initiating.<sup>2</sup>

In assessing pupil behavior, trained observers assigned each class of pupils observed a value, on a scale extending from 1 to 7, on each dimension. An assessment of 1 (at the left pole of the scale) on the Apathetic-Alert dimension indicated presumed observation of a preponderance of apathetic behaviors; an assessment of 7 (at the right pole) indicated presumed observation of many alert-type behaviors. Assessments were made for intact classes—for classes as a whole. A "Glossary," which accompanied the assessment form and which was employed in the training of observers, helped to standardize the assessment procedure.

Reliability coefficients for assessments of the several pupil behavior dimensions, based on correlation of independent assessments of first observer and second observer, ranged from .43 to .65 for some 1,500 elementary school classes and from .43 to .63 for approximately 1,900 secondary school classes.

Intercorrelations among the pupil behavior dimensions were substantial, ranging from .42 to .68 for secondary school classes and from .21 to .65 for elementary school classes. (Factorially, the pupil behavior dimensions for secondary school classes loaded on a single factor; for elementary school classes they contributed to two correlated factors, appearing to be roughly describable as pupil participation and controlled pupil activity.)

A minimum of two independent sets of assessments was obtained for each class (different observers observing the class at different times), the ratings of different observers subsequently being weighted equally (by transformation to a scale employing a common mean and standard deviation) and averaged to provide a composite assessment

<sup>2</sup> It may be noted that none of these pupil behavior dimensions directly reflects pupil learning—leastwise, not the learning of knowledge, understandings, and skills toward which much of the school's effort is directed. But most classroom practices do assume active attentiveness, purposeful and systematic study, reasoned self-assurance, and an initiating mind to be conditions of the pupil which are conducive to learning.

of each class of pupils on each of the bipolar dimensions considered.

A pupil behavior index,  $P_o$ , for each class of pupils was obtained by combining the averaged assessments on all four pupil behavior dimensions, equally weighted, into a single value. (Combination of the several dimensions into a single index seemed reasonable from a practical standpoint—as well as in light of their intercorrelations, which, as noted above, suggested a single secondary school cluster and two by no means distinct elementary clusters.) It was with this overall index,  $P_o$ , that the inventory estimated teacher characteristics subsequently were correlated.

The reliability of the pupil behavior index (the overall estimate contributed to by the four dimensions) was moderate. In one analysis (based on 98 classrooms) in which the pupil behavior dimension assessments of two observers were separately summed and the resulting  $P_o$  values correlated, the interobserver correlations were .62, .84, .72, and .64, respectively, when the observations were made approximately 30 minutes apart, more than 30 minutes apart but within the same half-day, during the same day but different half-day, and during the same school year but with an interval of more than one month.

### *Teacher Characteristics Data*

The Teacher Characteristics Schedule, employed to obtain estimates of teacher "traits" as described below, was an omnibus self-report type of inventory made up of items selected from a number of specially prepared separate instruments which had been subjected to a series of preliminary response-selection and validation studies. In its final form the schedule consisted of 300 multiple choice and check-list items relating to personal preferences, self-judgments, frequently engaged-in activities, biographical data, and the like. Ten scores (the last one, a score on a control variable used only to attempt to identify individuals with a strong tendency to make "socially acceptable" responses) were obtainable with the use of the Teacher Characteristics Schedule. Responses to the schedule had been selected and cross-validated (Ryans, 1960) to estimate teacher characteristics which were identified as:

warm, understanding, friendly vs. aloof, egocentric, restricted classroom behavior

responsible, businesslike, organized vs. evading, unplanned, slipshod classroom behavior

stimulating, imaginative vs. dull, routine classroom behavior

favorable vs. unfavorable opinions of pupils

favorable vs. unfavorable opinions of democratic classroom procedures

favorable vs. unfavorable opinions of administrative and other school personnel

learning centered ("traditional" or conservative) vs. child centered ("permissive" or liberal) educational viewpoints

superior verbal understanding (comprehension) vs. poor verbal understanding

emotional stability (adjustment or maturity) vs. instability

validity vs. invalidity of response (a control variable)

Generally, the reliability coefficients (stability and equivalence estimates) for scores derived from the Teacher Characteristics Schedule fell between .70 and .80. Validity coefficients (correlations between schedule scores and [a] observers' assessments of teacher classroom behavior, or [b] direct inquiry-type items relating to certain constructs or traits, e.g., attitudes, verbal ability, etc.) were of varying magnitude, depending upon the aspect of validity investigated (cross-validity, validity generalization, validity extension, concurrent validity, predictive validity), the particular teacher characteristic estimated, and the teacher sample (e.g., elementary teachers, secondary teachers, mathematics-science teachers, etc.) for which a particular scoring key was derived and to which it might most appropriately be applied. Concurrent validity coefficients typically ranged from .20 to .50 and cross-validity coefficients between .40 and .60. Coefficients of predictive validity were low, exceeding .20 only in two or three instances.

### *Sampling and Analysis*

Data from over 2,000 elementary and secondary school classes, widely scattered geographically, were collected and analyzed. For the most part the classes represented school systems of relatively large cities. Since cooperation of school systems, schools, and teachers was of necessity voluntary, the sampling could not comply with conditions theoretically required for comparison of the data with known probability models.

The teacher characteristics-pupil behavior data were analyzed separately for seven subsamples, three from elementary schools and four from secondary schools. The sizes of the subsamples varied, as shown in Table 1 below, from 99 to 718 classes of pupils and teachers. In each sample, and for each of the

TABLE 1

PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN "OBSERVER ASSESSED" PUPIL BEHAVIOR AND CERTAIN "INVENTORY ESTIMATED" TEACHER CHARACTERISTICS

"Inventory Estimated" Teacher Characteristic (with which "Observed" Pupil Behavior, $P_o$ , is compared)	Elementary School Samples <sup>a</sup>			Secondary School Samples <sup>b</sup>					
	A		B	D		E	F	G	H
	Female 718	Male 116	M & F 144	Female 510	Male 441	M & F 114	M & F 99	M & F 116	M & F 116
Sex of teacher: No. of classes:									
Understanding teacher behavior	.32*	.28*	.31*	.08	.11*	.07	.02	.10	
Businesslike teacher behavior	.36*	.29*	.31*	.15*	.11*	.00	.16	.22*	
Stimulating teacher behavior	.33*	.27*	.29*	.09*	.16*	.20*	.19*	.13	
Favorable attitude toward pupils	.18*	.15	.10	.03	.07	-.03	-.09	.14	
Favorable attitude toward democratic classroom	.18*	.14	.16*	.07	.12*	.02	-.08	.08	
Favorable attitude toward administrators and colleagues	.16*	.00	.15	.13*	.11*	.02	-.11	.06	
Conservative educational viewpoints <sup>c</sup>	-.14*	-.24*	-.09	.03	-.07	.01	-.14	-.15	
Verbal comprehension	.07	-.06	.11	.03	.10*	-.12	-.03	.32*	
Emotional stability	.13*	.04	.16*	.08	.11*	.14	-.09	-.02	
Validity of response	-.07	.00	.09	.02	.08	-.01	.02	-.11	

\* Asterisk used to note coefficients that differ significantly (.05 level or beyond) from .00.

<sup>a</sup> Samples A and C comprised of classes of Grades 1-6 pupils and their teachers; Sample B, of classes of Grades 3-6 pupils and their teachers.

<sup>b</sup> Samples D, E, F, and G comprised of classes of secondary school students and their teachers (roughly 22% mathematics classes, 25% science classes, 30% English classes, and 23% social studies classes); Sample H, of classes of foreign language students and their teachers.

<sup>c</sup> A high score on this characteristic reflects conservative, learning centered, educational viewpoints and a low score interpreted as indicative of a positive relationship between liberal educational viewpoints held by the teacher and productive pupil behavior in the teacher's class.

teacher characteristics studied,  $P_o$  of each class was paired with the teacher's score on the characteristic under consideration, and the product-moment correlation coefficient computed.

## RESULTS

*Elementary classes.* Data from the three elementary school samples suggest moderate, statistically significant (.05 level or beyond) interdependency relationships between assessments of observed purposeful and productive pupil behavior and inventory estimated understanding, friendly teacher behavior; organized, business-

like teacher behavior; and original, stimulating teacher behavior.

Although the remaining correlation coefficients are less uniformly significant from sample to sample, the evidence seems to point to a low positive relationship between pupil behavior and the inventory estimated teacher characteristics we have identified as favorable attitude toward democratic classroom procedures and liberal or child centered (permissive) educational viewpoints, and probably a similar low-order relationship between pupil behavior and favorable attitude toward pupils and emotional stability.

Samples A and C, involving female, and male and female teachers, respectively, both yield low positive correlations between pupil behavior and teachers' favorable attitude toward administrators and colleagues, but the coefficient for Sample B, male teachers, is of zero order.

With the notable exception of verbal intelligence all of the correlations between inventory estimates of teacher characteristics and observed pupil behavior were statistically significant in Sample A, the large sample of elementary classes taught by female teachers.

It seems likely that classroom pupil behavior is related to a number of teacher characteristics, of the kinds here estimated, when study is restricted to elementary school classes.

*Secondary classes.* The secondary school data suggest substantially fewer statistically significant relationships between the assessments of productive pupil behavior, as observed and assessed, and inventory estimated characteristics of teachers as measured by the Teacher Characteristics Schedule. It would appear that the only generalizable relationships are those represented by low correlations between observed pupil behavior and inventory estimates of original, stimulating teacher behavior—and, perhaps, between the pupil behavior index and the teacher characteristic interpreted as organized, businesslike classroom manner.

*Comparison of elementary and secondary school relationships.* As had been hypothesized from previously obtained evidence of correlations between assessments of directly observed pupil behavior and directly observed teacher behavior, interdependency relationships between pupil behavior and inventory estimated teacher characteristics were found to be distinctly

more apparent in elementary school classes and less discernible in secondary school classes.

For purposes of comparison of the pupil behavior-teacher characteristics correlations based on elementary vs. secondary school classes, consideration of the three elementary samples together and the five secondary samples together appeared justifiable. (For each of the inventory estimated teacher characteristics, application of the appropriate chi square test for the  $z_r$  transformations showed the values of the correlation coefficients between observed pupil behavior and the teacher characteristic under consideration to be no less homogeneous than might be expected from random sampling among the three elementary school samples and also among the five secondary school samples.)

When the pupil behavior-teacher characteristic correlation coefficients of the three elementary samples are translated into  $z_r$  values and combined to obtain the mean  $z_r$ , and this mean compared with the mean  $z_r$  value obtained by combining the five secondary school samples, the hypothesis that correlations between pupil behavior and teacher characteristics are more visible in the elementary as compared with the secondary school is supported—at least for the teacher characteristics referring to warm, understanding, friendly behavior; organized, businesslike behavior; stimulating, imaginative behavior; favorable opinions of pupils; favorable opinions of democratic classroom procedures; and child centered or permissive (liberal) educational viewpoints. The mean elementary  $z_r$  value was significantly higher (.05 level or beyond; two-tailed test) than the mean secondary  $z_r$  value for the inventory estimated teacher characteristics named. (Had the .10 level been

accepted as the criterion of significance, the pupil behavior-teacher emotional stability correlation also would have been significantly higher for the elementary compared with secondary school classes.)

It may be noted that tests of significance of the differences between  $z_r$  values for the elementary Sample A and secondary Sample D, both based on *female* teachers and their classes, also revealed significantly higher differences (.05 level or beyond) for the same teacher characteristics listed in the immediately preceding paragraph—with the pupil behavior-teacher emotional stability correlation also significantly higher in elementary classes if the .10 level of significance were employed. When *male* teachers and their classes were compared, elementary Sample B and secondary Sample E, *none* of the differences between  $z_r$  values attained the .05 significance level—although the differences between the correlations involving the understanding, friendly; responsible, businesslike; and child centered educational viewpoints on the part of teachers would have been significant had the .10 level been employed. Only two of the  $z_r$  differences—those involving understanding, friendly and responsible, businesslike teacher characteristics—met the .05 criterion when the relatively smaller independent samples of male and female teachers and their classes, Samples C and F, were compared—the difference of .23 between the elementary and secondary pupil behavior-teacher emotional stability  $z_r$  values being the next largest difference, but this significant only had the .10 level been accepted.

*Comparison of correlations involving male and female teachers.* As a final note, added in light of frequently expressed interest in the sex of the

teacher as a variable in the teacher-pupil relationship, it may be noted that the pupil behavior-teacher characteristics correlation coefficients obtained from samples of classes of female teachers and those obtained from samples of classes of male teachers generally are in the same direction, and although they vary slightly, such variation does not exceed that which might be attributed to random sampling. (When the correlations for Samples A and B, and similarly those for Samples D and E, were compared, none of the differences in correlation for men vs. women was significant—even at the .10 level.) The present data do not suggest systematic differences between men and women teachers insofar as the relationship between the teacher characteristics and pupil behavior estimated here is concerned.

#### SUMMARY

It was hypothesized that (a) certain teacher traits (characteristics) estimated from scores yielded by a self-report type inventory, the Teacher Characteristics Schedule, would be positively correlated with pupil behavior in the teachers' classes, and (b) pupil behavior-teacher characteristic relationships would be more in evidence in elementary than in secondary school classes.

Previously derived and cross-validated scoring keys were applied to teachers' responses to the Teacher Characteristics Schedule. Pupil behavior was directly observed and assessed, for intact classes, by trained observers.

Elementary school data suggest moderate relationships between assessments of observed purposeful and productive pupil behavior and inventory estimated teacher characteristics identified as "understanding, friendly behavior"; "organized, businesslike behavior"; and "original, stimulating

behavior"; and low-order relationships between pupil behavior and teachers' "favorable attitudes toward democratic classroom procedures," "child centered or permissive (liberal) educational viewpoints," "favorable attitudes toward pupils," and "emotional stability."

Secondary school data suggest substantially fewer relationships between the assessments of productive pupil behavior and inventory estimated teacher characteristics. The only generalizable relationships seem to be those between pupil behavior and inventory estimated "original, stimulating teacher behavior," and, possibly, between pupil behavior and inventory estimated "organized, businesslike teacher behavior."

Relationships between pupil behavior and inventory estimated teacher characteristics were less discernible in secondary school classes as compared with elementary school classes.

No significant differences were observed between the pupil behavior-teacher characteristic correlations of men teachers as compared with women teachers.

#### REFERENCES

- RYANS, D. G. *Characteristics of teachers*. Washington, D. C.: American Council on Education, 1960.
- RYANS, D. G. Some relationships between pupil behavior and certain teacher characteristics. *J. educ. Psychol.*, 1961, **52**, 82-90.

(Received September 6, 1960)

# RELATIONSHIP OF INTELLIGENCE TO STEP SIZE ON A TEACHING MACHINE PROGRAM<sup>1</sup>

CARLETON B. SHAY<sup>2</sup>

*University of California, Los Angeles*

Teaching machine programs have generally been written with little consideration for different levels of intellectual ability. The programmed teaching method itself has been considered sufficient to overcome ability differences, because most programs have been written with minimal step size, allowing progression through the programs with little error and maximum reinforcement. Skinner (1958) suggests

a program designed for the slowest student in the school system will probably not seriously delay the fast student, who will be free to progress at his own speed... (p. 976).

On the other hand, since the intellectually superior student should be able to think in larger steps than one less intelligent, it seems likely there is a relationship between intelligence and step size. Experimental evidence on such a relationship is meager: Porter's (1959) study with spelling words, which found no relationship between IQ and achievement, supports the contention that one program is sufficient for different ability levels; but results of a study by Briggs and Besnard (1956), using more complex materials, suggest that programs suited to different ability levels might be desirable.

The present study investigated the hypothesis, in null form, that there is no relationship between intelligence and step size on a teaching machine program for each of the following cri-

<sup>1</sup>This article is based on the author's doctoral dissertation at the University of California, Los Angeles, under the direction of E. R. Keislar.

<sup>2</sup>Now with the Santa Monica, California, city schools.

teria: total learning, learning of "rote" materials, learning of materials involving "understanding," errors, and time to complete the program.

Size of item step was defined, after Lumsdaine (1959), as the "difficulty of giving the correct answer," and inferred from measurement of the number of errors made on a program. The proportion of errors on a program gives a difficulty level from which the average probability of correct response may be calculated thus:

$$p\{R_j\} = 1 - \frac{\sum_i \sum_{i,j} E_{ij}}{N_i N_j}$$

where

$p\{R_j\}$  = average probability of correct response  $R$  on items  $j$ ;

$E_{ij}$  = number of errors made by the  $i^{\text{th}}$  individual on the  $j^{\text{th}}$  item;

$N_i$  = sum of the  $i$  individuals; and

$N_j$  = sum of the  $j$  items.

The probability of responding correctly to any given item is:

$$p\{R_j\} = 1 - \frac{\sum_i E_{ij}}{N_i}$$

Both the criteria of individual and average item probabilities of correct response were used to adjust and evaluate step size in the programs used in this study.

## METHOD

*Programs* Three programs were written to cover the fourth grade unit in roman numerals.<sup>3</sup>

<sup>3</sup>The programs are available with the author's dissertation (Shay, 1960) through interlibrary loan. They are under separate cover, however, and should be requested specifically.

This unit presents the symbols and principles for construction of roman numerals to 399 (Brueckner, Merton, & Grossnickle, 1957). The programs with few exceptions, however, did not venture beyond 100.

A large step program was written first and given to successive small ( $N \sim 4$ ) samples of above average ( $IQ \geq 110$ ) pupils, each sample followed by item revision until the average probability of correct response was  $\geq 90\%$ , and most individual item probabilities of correct response were  $\geq 80\%$ . This large step program was similarly reduced in step size for samples of average students ( $IQ 90-109$ ) to form a medium step program; this, in turn, was revised to form a small step program for below average pupils ( $IQ \leq 89$ ). Step size was reduced by the insertion of bridging items, splitting an item into two or more simpler ones, or by rewording stems or responses to provide more support to the slower student.

The three programs were next administered to a sample of 90 students, an equal number at each ability level and with each program. The step size was then readjusted to form the final experimental programs. They contained 103 (large step), 150 (medium step), and 199 (small step) items, and were duplicated and ring-bound to form small booklets approximately 4" x 6" in size. The first three items in each program were used to instruct the response procedures and were not considered in the error count or timing. With decreasing step size, programs were less difficult and contained more review items and more varied examples illustrating the principles of roman numerals construction.

### Apparatus

Twelve identical response devices were used. Each device consisted of a board faced with aluminum foil upon which was clamped a standard IBM answer sheet and an IBM matrix (key) faced with aluminum foil and punched for the correct answers. The student responded by punching through the answer sheet with a stylus, completing an electrical circuit to either a green or red light, informing the subject of his success.

### Subjects

Ninety subjects were chosen from the low fourth grade in four Los Angeles elementary schools on the basis of roman numerals pretest and group intelligence test scores. One additional subject was dropped because of absence, and a second because of

complete failure to follow directions. A third of the subjects were above average in intelligence ( $IQ \geq 110$ ), one-third average ( $IQ 93-109$ ), and one-third below average ( $IQ \leq 92$ ). Adjustment of the customary lower limit of the average range was necessitated by a lack of acceptable low ability students.

### Pretest and Posttest

The pretest consisted of 24 items, 12 to be written in roman numerals from arabic, and 12 vice versa. They were all numbers from 1-39 except for L, C, D, M, 40, and 44. Only the number 8 was common to both lists.

The posttest consisted of 31 items: 10 arabic numbers to be written in roman numerals, 14 of the opposite type, and 7 from a sequence which required counting by 100s to 1,000 in roman numerals, with C, D, and M given. The first two parts mentioned were arranged with old items first, followed by new numbers not taught in the programs. The old items were those numerals specifically taught and presumably not requiring an understanding of roman numerals construction for recall. The term "rote," as applied to these items, refers only to the minimal amount of understanding necessary for learning and not to the method of presentation. In contrast, the new items had not been taught and could be formulated only through an understanding of these principles. Most of the new items involved inversions (as in 4 and 9) and were numbers between 100 and 1,000. There were 12 old and 19 new items. Reliability of the posttest, as estimated by the application of Kuder-Richardson Formula 20 to the scores of the experimental subjects, was as follows: total posttest, .93; old items portion, .83; and new items portion, .90.

### Procedure

The pretest was administered to all students in each 4B class. Group test IQs were taken from the records of students whose pretest scores were from 4 to 20, and subjects assigned to ability groups. Subjects within each ability group were randomly assigned to one of the three programs to form nine experimental groups of 10 subjects each. These groups, explicitly defined in Table 4, were formed from each possible pairing of ability group and program.

Subjects worked on the programs in groups of 7 to 12, typically including members from 5 or 6 of the experimental groups. In the first experimental session for each group, subjects were instructed in the re-

sponse procedures with the first three items. Each subject completed 50 items a day on consecutive school days until he finished his program; this was followed immediately by the posttest. Performance measures obtained for each subject included old and new items posttest and total posttest scores, total errors, and time necessary to complete the program.

## RESULTS

Tables 1 and 2 indicate the extent to which the programs met the criteria of average and individual item difficulties for the experimental groups.

Values in the lower-left to upper-right hand diagonal in each table, representing the criterion ability

group-program combinations, suggest that the programs were adequate for the study. Additional examination of the location of items within each program not meeting the individual item difficulty criterion revealed a slight tendency in the small step program for these items to cluster. This trend was not considered detrimental to the subjects' responses or to the further interpretation of data.

Desirability of covariance analyses of criterion scores with pretest scores as a control variable was indicated by the following correlations: pretest-IQ, .33; pretest-total posttest, .66; pretest-new items posttest, .59; pretest-old items posttest, .66; pretest-errors, -.51; and pretest-time, -.11.

Covariance analyses of criterion scores are summarized in Table 3. The assumption of homogeneity of regression was tested in each case and found to be satisfied. Inverse sine transformations of percentage error scores, and logarithmic transformations of time scores were used to satisfy the assumption of homogeneity of variance. The hypothesis that there is no relationship between intelligence and size of item step could be rejected only in the case of percentage error, where an interaction was found ( $p < .05$ ). The main effect of ability level was obtained for each of the other criterion scores, but no interaction.

Adjusted mean criterion scores for each experimental group on each of the variables—new, old, and total posttest, time, and errors—are given in Table 4.

## DISCUSSION

Failure to reject most of the null hypotheses asserted in this study supports Skinner's position that it is not necessary to provide more than one program on the basis of different initial ability. Table 4 suggests that at all

TABLE 1  
MEAN NUMBER OF ERRORS AND MEAN PROBABILITY OF CORRECT RESPONSE

Program	Ability Group		
	Below Average	Average	Above Average
Large step			
$M$	36.6	18.9	9.4
$\bar{p}\{R_i\}$	63.4	81.1	90.6
Medium step			
$M$	34.0	15.3	15.0
$\bar{p}\{R_i\}$	76.9	89.6	89.8
Small step			
$M$	22.1	25.6	17.2
$\bar{p}\{R_i\}$	88.7	86.9	91.2

TABLE 2  
NUMBER AND PERCENTAGE OF ITEMS WHOSE CORRECT RESPONSE PROBABILITY EQUALS OR EXCEEDS 80%

Program	Ability Level		
	Below Average	Average	Above Average
Large step			
$N$	40	73	90
%	40.0	73.0	90.0
Medium step			
$N$	92	131	137
%	62.6	89.1	93.2
Small step			
$N$	171	174	186
%	87.2	88.8	94.9

TABLE 3  
COVARIANCE ANALYSIS OF CRITERION SCORES USING A ROMAN NUMERALS  
PRETEST AS A CONTROL

Source	df	Mean Square					F				
		Total posttest	New items posttest	Old items posttest	$\sin^{-1}$ percentage errors	Log time	Total posttest	New items posttest	Old items posttest	$\sin^{-1}$ percentage errors	Log time
Programs	2	66.18	28.21	8.13	501.50	1.93	1.80	1.45	1.30		27.21**
Ability	2	382.61	160.01	48.56	480.25	1.36	10.43**	8.23**	7.78**	3.17*	19.17**
Interaction	4	68.62	26.15	14.08	189.56	.02	1.87	1.35	2.26		.22
Within cells	80	36.70	19.44	6.24	59.88	.07					

\* .05 level of significance.

\*\* .001 level of significance.

TABLE 4  
ADJUSTED MEAN CRITERION SCORES AND STANDARD DEVIATIONS

Experimental group		Total posttest		New items posttest		Old items posttest		Time (minutes)		Percentage error	
		M	SD	M	SD	M	SD	M	SD	M	SD
Above Average	Ability	Large step	19.2	5.7	10.3	4.9	8.8	1.3	49.0	12.7	10.5
		Medium step	19.9	7.7	10.8	5.9	9.1	2.4	76.1	15.7	9.2
		Small step	20.9	7.1	12.5	4.4	8.5	2.9	90.0	19.3	7.8
Average	Ability	Large step	12.9	7.4	6.5	5.0	6.4	3.4	75.6	14.7	17.7
		Medium step	18.3	6.2	10.0	4.8	8.3	1.9	90.6	17.5	9.0
		Small step	15.8	8.2	8.2	6.0	7.6	3.2	122.0	22.7	13.0
Below Average	Ability	Large step	12.1	8.2	6.5	4.8	5.6	3.8	87.3	32.9	32.5
		Medium step	9.4	7.0	4.5	3.2	4.9	3.9	108.5	35.1	18.5
		Small step	16.4	8.3	8.4	5.3	8.0	3.5	135.5	42.5	9.3

levels of ability and for different types of material, as size of item step decreases, there is an accompanying increase in posttest score and time to complete the program.

The question naturally arises, why did the experiment fail to reveal a relationship between intelligence and size of step, assuming there is a relationship? One possible reason is that the programs were inadequate. Perhaps a 90% probability of obtaining the correct response is too low. This is suggested by the inverse relationship found between posttest score and step

size. An argument against this explanation is the possibility that extraneous factors produced a portion of the error, and the "true" error rate was actually lower. Machine malfunction was known to cause some error, which varied according to the subject's reaction to occasionally seeing both lights at once. Also, some subjects exhibited a certain confusion between an answer and the number of the answer. For example, if the question was "What is 5 in roman numerals?" and the answer V appeared at Choice Position 3, some subjects would answer at

Choice Position 5 rather than 3. (In future experiments, lettered response positions are recommended.)

A second possible reason for not finding a relationship between intelligence and step size is that IQ did not serve as an adequate predictor of learning ability in this study. Studies generally show only a moderate relationship between learning ability and IQ. Perhaps group intelligence tests are less effective predictors of learning ability where programmed teaching sequences are involved than is the case in the ordinary classroom. Since programs are carefully regulated in terms of step size, reading level, sequencing, etc., it may be that general intelligence as it is usually measured plays a smaller role in this type of learning than in the classroom where these variables are not controlled. An implication of this is that other, perhaps more specific measures of ability should be used to predict learning from programmed teaching.

A third explanation is the possibility that a "real" relationship was obscured by the confounding of certain pretest relationships and the experimental design. Considering the correlations between pretest-IQ and pretest-posttest, and assuming that IQ also contributed to the variance in posttest scores, the use of pretest scores as a control variable reduced the contribution of IQ to the variance of the criterion variables. An implication of this explanation is that to test the hypothesis adequately, it would be necessary to use some material such as an artificial language, whose pretest scores would be essentially equal and unrelated to IQ. A further reason for not obtaining significant interactions is the possibility that 10 cases per cell were too few to reveal a relationship.

Another explanation deals with the definition of step size. If step size had

been defined in terms of the amount of language or prior experience necessary to make the transition from one step to the next, it is probably more likely that interactions would have been found.

Now let us assume that in fact there is no relationship between ability and step size. This implies, as Skinner suggests, that for programmed teaching a minimally-small step program is the most appropriate for all levels of ability. This would further suggest that if provision for individual differences is to be made, it should be made on some basis other than intelligence alone. Such a provision is indicated by the time data, perhaps. Comparing large and small step programs, bright students took almost twice the time to improve their score 1.7 points, an 8.9% increase. In some situations, such a gain may not be worth the additional expenditure of time, and the use of more than one program would be indicated.

It would be desirable to explore the hypothesis of this study further, with the following differences: more subjects; easier programs; different learning material, such as an artificial language; different age groups; and use of different predictors of "learning ability" from general to task-specific.

#### SUMMARY

This study examined the hypothesis that there is no relationship between intelligence and size of item step on a teaching machine program under the criterion conditions of total learning, learning involving rote materials, learning of materials involving understanding, errors, and time to complete the program. Step size was defined as the difficulty of giving the correct answer and measured by two criteria

Three programmed teaching sequences of 103, 150, and 199 items were developed covering fourth grade Roman numerals. Each was written for a given level of ability, and step size adjusted to meet stated criteria. Ninety fourth graders were selected on the basis of pretest and intelligence test scores. From each of three ability levels, three groups of 10 subjects were randomly selected and each group assigned to one of the programs, making nine experimental groups. Subjects completed 50 items a day on successive days until program completion, which was immediately followed by the posttest.

None of the tested null hypotheses could be rejected except in the case of percentage error score. The results, within the definition of step size used, indicate that if there is a relationship between intelligence and step size, it is not a strong one. This would suggest that alternate programs are not necessary on the basis of ability alone. Further studies were suggested.

## REFERENCES

- BRIGGS, L. J., & BESNARD, G. Experimental procedures for increasing reinforced practice in training air force mechanics for an electronic system. In G. Finch & F. Cameron (Eds.), *Symposium on air force human engineering, personnel, and training research. NAS-NRC Publ.*, 1956, No. 455. Pp. 48-58.
- BRUECKNER, L., MERTON, E., & GROSSNICKLE, F. *The new learning numbers*. New York: Winston, 1957.
- LUMSDAINE, A. A. The development of teaching machines and programmed self-instruction. Paper read at Conference on Status of Research and Experimentation on the Newer Instructional Media. Center for Advanced Study in the Behavioral Sciences, Stanford, November 1959.
- PORTER, D. Some effects of year long teaching machine instruction. In E. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959. Pp. 85-90.
- SHAY, C. B. The relationship of intelligence to size of item step on a teaching machine program. Unpublished doctoral dissertation, University of California, Los Angeles, 1960.
- SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969-977.

(Received August 9, 1960)

## TRANSFER OF VERBAL MATERIAL ACROSS SENSE MODALITIES

PAUL PIMSLEUR AND ROBERT J. BONKOWSKI<sup>1</sup>

*University of California, Los Angeles*

One of the main arguments against the teaching of the spoken foreign language in high schools and colleges maintains that there is not enough time in the average course in which to teach both speaking and reading. However, some writers contend that students who learn first to speak the language will, within the normal course time, catch up to or perhaps even surpass in reading ability those who have been taught reading all along. The argument of those who espouse this view will be bolstered if it can be demonstrated experimentally that aural learning facilitates visual learning.

Experiments bearing directly on the transfer of verbal material across the modalities of vision and audition are rather scarce in the literature of experimental psychology. Weissman and Crockett (1957) demonstrated that transfer does occur from auditory training to visual discrimination. Postman and Rosenzweig (1956) suggested that transfer from visual training to auditory discrimination is greater than conversely. However, in both of the above studies the investigators were concerned with the thresholds of recognition of verbal material.

The present study addressed itself to several questions raised by the controversy over the teaching of languages: (a) Does aural learning facilitate visual learning? (b) Is this facilitation greater than that achieved by presenting the material first visually

and then aurally? (c) In terms of total time, is it more economical to teach by the aural-visual order than by the visual-aural order (i.e., does the student take fewer total trials to learn verbal material both visually and aurally when the material is presented first aurally and then visually)? Affirmative answers to these questions would lend some support to those favoring the aural approach to language teaching.

### METHOD

*Subjects.* The subjects were 28 paid volunteers recruited from among students enrolled in second semester Spanish courses (Spanish II) at UCLA. The group consisted of 10 females and 18 males.

*Design.* A treatments  $\times$  levels design (Lindquist, 1956, pp. 121-149) was used. The subjects were first divided into two levels: those who had received a grade of A or B in Spanish I at UCLA (14); and those who had received a grade of C or D in Spanish I at UCLA (14). At each level, subjects were randomly assigned to one of two conditions by the use of a table of random numbers.

*Procedure.* In the present design there were two treatments. The subjects learned a list of paired associates through one modality, and relearned the same list through another modality. Original learning and relearning were continued to a criterion of two consecutive errorless trials. Group V-A learned first through a visual presentation of the list and relearned through an aural presentation of the list. Group A-V received the opposite order of presentation of the same list: aural original learning and visual relearning. For the purposes of the present study, these conditions enabled each group to serve as a control for the other. The original learning scores of one group could be compared with the relearning scores of the other group as in a simple transfer paradigm.

The list consisted of 10 nonsense dissyl-

<sup>1</sup>The authors wish to express their gratitude to Norman H. Anderson for his helpful suggestions and comments on a draft of this paper.

ables, each paired with one of 10 color names. The dissyllables, from Dunlap's (1933) list, were chosen to meet requirements of ease of spelling, ease of recognition aurally and visually, and similarity in construction and pronounceability to English words. The stimulus-response pairs are listed in Table 1.

In the visual presentation the stimulus word was projected on a screen for 1 second and then removed. After a 4-second interval the response word was projected for 1 second. There was a 4-second interval between pairs. In the aural presentation the pairs were presented by tape recorder, with a 4-second interval both between stimulus and response and between pairs. Since a word took about 1 second to pronounce, the time intervals were comparable.

Using the anticipation method in each condition, the subject responded by writing his response on a line of a record sheet using a strip of cardboard to cover previous responses. The 10 responses in each trial were recorded on successive sheets in a booklet. The 10 pairs were presented in different random orders for successive trials in a condition, but the orders were the same between conditions.

The subjects were run in small groups ranging from two to four in order that the experimenters could maintain control and prevent cheating. The subjects were instructed to leave the room quietly after having reached criterion in order to prevent additional learning by subjects who finished first. The relearning phase of a condition was begun as soon as all subjects being run at one time had reached criterion in the original learning phase.

The subjects were instructed to learn the "meanings of 10 foreign words." Virtually the same instructions were given before original learning and relearning so that subjects did not know in advance that the "10 foreign words" and their "meanings" in the relearning phase were the same as in the original learning phase.

#### RESULTS AND DISCUSSION

The dependent variables were trials and errors to criterion. The correlations between these two measures were .89 for the original learning phase and .90 for the relearning phase. Statistical tests based on both measures yielded similar results. Therefore, only the data for number of trials are reported.

TABLE 1  
TEN PAIRS OF DISSYLLABLES  
AND COLOR NAMES

polef	green	fimur	yellow
medon	purple	runil	brown
defig	red	kupod	white
nigat	pink	tarup	black
gokem	blue	latuk	orange

TABLE 2  
MEAN TRIALS TO CRITERION IN ORIGINAL  
LEARNING AND RELEARNING

Spanish I Grade	Original Learning		Relearning	
	Group V-A	Group A-V	Group V-A	Group A-V
A or B	8.86	6.86	2.71	2.29
C or D	14.57	9.14	4.14	2.57
Group Mean	11.71	8.00	3.43	2.43
Group SD	4.67	2.78	2.87	.62

The data for the analysis of amount of transfer are summarized in Table 2. Comparing the relearning scores for each group with the original learning scores of the other group suggested that there was a marked positive transfer across modalities. Aural original learning facilitated visual relearning and the converse was also true. The facilitation of visual learning by prior aural learning may be viewed as a finding in favor of the aural approach to language teaching.

Stevens' (1951, p. 557) formula for the savings score was applied in the analysis of the differential degree of transfer between the two conditions. The mean savings scores for Groups V-A and A-V were 89.5 and 94.6, respectively. This showed a tendency for subjects in the A-V condition to "save more" than subjects in the V-A condition. However, the savings score is to some extent dependent upon the criterion set by the experimenter. Therefore, an attempt was made to get a

TABLE 3  
SUMMARY OF ANALYSES OF VARIANCE OF  
ORIGINAL LEARNING AND OVERALL DATA

Source	df	Original Learning		Overall (OL + RL)	
		MS	F	MS	F
Sense Modalities	1	96.59	8.27**	155.64	5.86*
Grades	1	112.01	9.59**	165.20	6.22*
SM × G	1	20.56	1.76	36.65	1.38
Error	24	11.68		26.56	

\* Significant at .05 level.

\*\* Significant at .01 level.

more meaningful analysis of the differential degree of transfer.

In the original learning phase subjects at both levels of Group V-A appeared to have taken more trials to criterion, on the average than their counterparts in Group A-V. Similarly, looking at the performance of subjects in each group suggested that the overall mean for Group V-A was greater than that of Group A-V in the original learning phase. An analysis of variance was employed to test the statistical significance of these differences. A summary of this analysis appears in Table 3. The significant *F* for Sense Modalities suggests that the visual learning was more difficult than the aural learning in the original phase. This finding indirectly supports the conclusion that there was more transfer in the A-V condition than in the V-A condition. For if one can agree that the visual task was in fact more difficult, one would not expect subjects in the A-V condition to take a smaller mean number of trials to criterion in the relearning phase. Nevertheless, subjects in the A-V condition obtained a mean of 2.43 trials for the relearning phase as opposed to a mean of 3.43 trials for subjects in the V-A condition. It should be stressed that this is offered only as indirect support of the

conclusion that there was greater transfer in the A-V than in the V-A condition.

The most significant finding, and perhaps the strongest support for those advocates of the aural approach to language teaching, resulted from the analysis of the overall (OL + RL) scores. Combining the number of trials for each subject in both phases of the experiment yielded scores that gave some indication of the effectiveness of the *order* of presentation of verbal material. A summary of the analysis of variance of these data is given in the last column of Table 3.

The significant *F* for Sense Modalities in this case suggests that subjects in the A-V condition took on the average fewer trials to both criteria than subjects in the V-A condition. Hence, it took the students fewer total trials to learn the verbal material both visually and aurally when the material was presented first aurally and then visually. This has the implication that it might be more economical, in terms of time, to teach verbal material by the aural-visual order than by the visual-aural order.

The significant *F* for Grades both in the original learning and the overall data suggested that those subjects who achieved a grade of A or B in Spanish I took fewer trials to criterion, on the average, than those subjects who achieved a grade of C or D. In itself, such a result might have been predicted by the experienced language teacher. However, the Grades effect is of importance also in considering the methodology of transfer studies similar to the present study. The design used afforded a neat method of matching the groups as well as increasing the efficiency of the experiment. In fact, had the groups not been so matched, the estimate of the error mean square would have been in-

creased. The inflated error mean square would have reduced the chances of obtaining a significant Sense Modalities effect.

On the basis of the results of this study the investigators concluded: aural learning does facilitate visual relearning, this facilitation is greater than the facilitation of aural relearning by visual original learning, and it takes less time to teach verbal material both visually and aurally when the material is presented first aurally and then visually.

### SUMMARY

A list of 10 paired associates (disyllables as stimuli and color names as responses) was randomly presented first through one modality and then through another modality. Half the subjects learned the list first through the visual modality and then relearned it through the auditory modality. The other half learned the list in the opposite order.

Positive transfer was found in both directions. It was suggested that the aural presentation had a greater facil-

tating effect upon the visual presentation than conversely. The subjects took fewer total trials to learn verbal material both visually and aurally when the material was presented first aurally and then visually. These findings seemingly offer some support for the view that aural instruction preceding visual instruction may have advantage over conventional methods of language teaching if the goal is to achieve proficiency in both reading and aural comprehension.

### REFERENCES

- DUNLAP, K. *List of 43,200 dissyllable words and paralogs*. Washington, D. C.: National Research Council, 1933.
- LINDQUIST, E. V. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
- POSTMAN, L., & ROSENZWEIG, M. R. Practice and transfer in the visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1956, **69**, 209-226.
- STEVENS, S. S. *Handbook of experimental psychology*. New York: Wiley, 1951.
- WEISSMAN, S. L., & CROCKETT, W. H. Inter-sensory transfer in the visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1957, **70**, 283-285.

(Received August 29, 1960)

# VERBAL CONCEPT LEARNING IN HIGH SCHOOL STUDENTS WITH PICTORIAL AND VERBAL REPRESENTATION OF STIMULI

WILLARD N. RUNQUIST AND VALORIE H. HUTT

*Pomona College*

Within an association framework, concept formation has been conceived of as the process by which a group of stimuli come to consistently elicit the same response. Obviously, one important factor determining the rate of concept formation, then, would be the ease with which the stimuli elicit the common response.

In order to study the formation of verbal concepts, Underwood and Richardson (1956a) scaled a series of common nouns in terms of their dominant descriptive associations. Concept materials were then formed by selecting groups of nouns having the same dominant association. These were presented to subjects who had to discover and learn the correct associations. Further studies (Freedman & Mednick, 1958; Underwood, 1957; Underwood & Richardson, 1956b) utilizing these materials have explored several variables affecting the rate of concept learning.

In all of the studies cited above, the stimuli (the nouns) were all presented verbally (i.e., in terms of the word representing the object). It has been suggested (Underwood, 1952) that other modes of presentation might have some effect on the rate of concept learning of the type described above. The purpose of the present study was to compare the learning of verbal concepts with pictorial representation of stimuli with the more standard verbal representation.

It would seem that a case might be made for the superiority of either condition. Verbal stimuli might be easier

because irrelevant parts of the pictures could interfere with the production or discovery of the correct responses. On the other hand, with pictorial representation, the correct association (a descriptive adjective) could be made quite apparent from the nature of the picture, while in the case of the verbal representation, the subject is free to "picture" the object any number of possible ways, and may picture it in such a way as to interfere with the discovery of the correct response.

In order to provide data relevant to these hypotheses, three conditions were utilized in the present study: Verbal (the name of the object), Picture Dominant (with the correct association emphasized by the picture), and Picture Nondominant (with the correct association de-emphasized). Grade level was also incorporated as one variable in the design, as it seemed desirable to be able to extend and generalize any results to include several educational levels if possible.

## METHOD

*Materials.* The concepts and stimuli were chosen directly from Underwood and Richardson's materials (1956a). These are presented in Table 1. Certain criteria were used in selecting concepts. Only high dominat concepts were used, overlap and intra-list similarity were kept as low as possible, and all nouns had to be easily represented in pictorial form.

Three sets of the stimuli were hand drawn on 5" × 6" railroad board cards in black india ink. One set consisted of the words representing each noun. The second and third sets were line drawings of the objects.

In one set, the descriptive characteristic which was to be the correct response was emphasized by the drawing (Picture Dominant). In the other set, this was de-emphasized (Picture Nondominant). For example, a bed was drawn as appearing soft and billowy for one set, and looking like an army cot for the other.

**Subjects.** Sixty high school students (15 at each of four grade levels: ninth, tenth, eleventh, and twelfth) served as subjects in this study.<sup>1</sup> Three groups of 20 subjects each were formed with 5 subjects from each grade level in each of the three groups. No attempt was made to match groups, with subjects being assigned in order of appearance. The IQ level of all students ranged from 90-110. Each group learned one set of the concept materials.

**Procedure.** Each subject was run individually, seated at a table facing the experimenter. Stimuli were exposed manually for 3 seconds with a 2-second interval between them. Intervals were timed by an electronic timer with an audible click. Instructions were essentially the same as those used by Underwood and Richardson (1956b). The subject was told to respond during the 3-second interval with descriptive words, i.e., those expressing size, shape, texture, etc. The instructions were not completely standardized as it was necessary that each subject understood the task. One example (Elephant-big) was used in which examples of descriptive words were given. The subjects were also told that only one "response" was correct and that the whole list took only four responses. The experimenter said "right" after all correct responses and said nothing if the subject had not responded correctly within 3 seconds. Each subject received 15 repetitions of the list, with the cards being shuffled after each repetition.

## RESULTS

The mean number of correct responses for each group over all 15 trials is shown in Figure 1. Analysis of variance of these data showed that both mode of presentation and grade level were significant ( $F$  for mode of presentation was 15.26; for grade level 5.87). Individual  $t$  tests showed the verbal instances produced better per-

<sup>1</sup> Subjects for this study were provided by Azusa High School, Azusa, California; Nelson Price, Principal.

TABLE 1  
CONCEPT MATERIALS USED

Concept	Instances
round	baseball doughnut barrel spool
soft	bed fur pillow moccasin
sharp	fang fishhook hatchet knife
slimy	lizard earthworm oyster snake

formance than the picture emphasized instances ( $t = 2.67, p < .01$ ) and the picture emphasized instances better performance than the pictures de-emphasized ( $t = 2.54, p < .02$ ). As to grade level, the only significant difference was between Grades 11 and 12. There was no interaction between mode of presentation and grade level ( $F = 1.29$ ), even though the graph suggests that the differences were larger at the lower grade levels.

## DISCUSSION

The most important finding of this study is that performance in concept learning was better when the stimuli were in the form of words than when they were represented pictorially. There are several interpretations, two of which seem likely. First, the subject is using the same medium as the stimulus in making his response, i.e., he is looking at a word and responding with a word. Since the words used here are common ones, it is reasonable to assume that the concept and stimuli

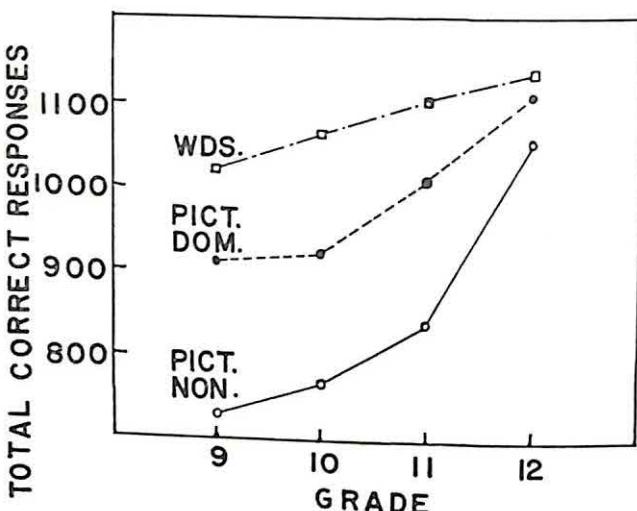


FIG. 1. Mean correct responses as a function of grade level for Verbal, Picture Dominant, and Picture Nondominant instances.

have been associated during the subject's prior experience. Hence, in emitting the correct response, the subject may not "form an image" at all, but merely responds with a highly likely verbal associate. Karwoski, Gramlich, and Arnott (1944) came to somewhat the same conclusion from their data which show a longer reaction time to objects than to words.

The interpretation which seems most plausible, however, attributes these findings to the particular concepts which were used. Such qualities as "soft," "sharp," and "slimy" are primarily tactal rather than visual. It is very likely, then, that pictorial representation of these objects suggested visual associations such as color, form, etc., or at least interfered with the production of an association of a nonvisual nature. One test of this might have been to determine whether the "wrong" associations or "first" associations given to the instances of these concepts were of this nature. However, the mechanics of administering the task precluded recording these responses. Certainly, the hypotheses advanced here would seemingly predict that Underwood-Rich-

ardson Dominance Values scaled to the pictures would be different from those obtained with words. Even this, however, would not be crucial since the Dominance Values do not necessarily represent the relative strengths of responses within a single individual, but merely the percentage of the subjects who gave a particular response as the first association. Interfering response tendencies might not be manifested in these Dominance Values, but could appear as lengthened response time.

This interpretation suggests that broad generalizations regarding the relative effectiveness of pictures, objects, and words as materials for concept formation will in all probability not be found. As is the case in many complex learning situations, the effects of a particular variable will probably be found to depend upon the specific nature of concepts to be learned.

Mention perhaps should be made of the fact that these results are contrary to those obtained in a paired-associate task by Wimer and Lambert (1959). They found learning to be superior to object stimuli as opposed to word stimuli and attribute these results to

the greater similarity of the verbal stimuli. In reconciling the present findings with these results, two things might be pointed out: First, the similarity of the materials used here might differ considerably from that in the materials used by Wimer and Lambert who used actual objects, not pictures. Second, similarity does not act the same in concept formation studies as in paired-associate learning and, in fact, any similarity between instances of the same concept should actually facilitate acquisition (Richardson, 1958). Thus, there is probably no incompatibility of the present findings with those of Wimer and Lambert.

The fact that the more advanced subjects performed better on this task probably may be attributed to increasing mental age. There is, in addition, some suggestion that the differences between methods of presentation are not as marked in the twelfth graders as in the other groups. Although, statistically speaking, the conservative conclusion is that the trend is unreliable, its regularity suggests a real phenomenon. It could have been produced by something in the nature of the associations elicited by the stimuli or perhaps by a more sophisticated approach on the part of the more advanced students. At any rate, any explanation for this somewhat questionable phenomenon must await further exploration.

#### SUMMARY

The formation of verbal concepts was investigated in high school students utilizing instances consisting of words, pictures of objects with the correct concept accentuated, or pictures

of objects with the correct concept de-emphasized. Four different grade levels were combined with the three modes of presentation of the concepts in a factorial design. The results showed that concept learning was best when the instances were presented as words, followed by the picture stimuli with the concept accentuated. The results were attributed to the nature of the specific concepts studied, and it was concluded that in all likelihood no simple answer to the question of which type of presentation of instances is superior can be obtained.

#### REFERENCES

- FREEDMAN, J. L., & MEDNICK, S. A. Ease of attainment of concepts as a function of response dominance variance. *J. exp. Psychol.*, 1958, **55**, 463-466.
- KARWOSKI, T., GRAMLICH, F. W., & ARNOTT, P. Psychological studies in semantics: I. Free association reactions to words, drawings, and objects. *J. soc. Psychol.*, 1944, **20**, 233-247.
- RICHARDSON, J. The relationship of stimulus similarity and number of responses. *J. exp. Psychol.*, 1958, **56**, 478-484.
- UNDERWOOD, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **89**, 209-220.
- UNDERWOOD, B. J. Studies of distributed practice: XV. Verbal concept learning as a function of intralist interference. *J. exp. Psychol.*, 1957, **54**, 33-40.
- UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84-95. (a)
- UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229-238. (b)
- WIMER, C. C., & LAMBERT, W. L. The differential effects of word and object stimuli on the learning of paired associates. *J. exp. Psychol.*, 1959, **57**, 31-36.

(Received September 26, 1960)

## GRADES AS REINFORCING CONTINGENCIES AND ATTITUDE CHANGE

ROBERT N. BOSTROM,<sup>1</sup> JOHN W. VLANDIS,<sup>2</sup> AND MILTON E. ROSENBAUM  
*State University of Iowa*

Although the use of marks and grades in educational practice has received considerable attention in the research literature, little if any information is available about the effects of grades on the students to whom they are given (Odell, 1950). The present study was designed to examine the effect on attitudes of differential assignment of grades for performance on attitude related essays.

The basic hypothesis guiding the study was that grades will serve to affect behavior in the form of a reinforcing contingency. More specifically, a "good" grade should serve to effect repetition of the responses which it followed, while a "poor" grade should produce reduction in the potentiality of appearance of the preceding responses.

In the present study, students responded to attitude scales and several weeks later were requested to write essays on topics related to these attitudes, adopting a position incongruent with their measured attitudes. At random, grades were assigned to these essays. It was predicted that students who received an A for their essays should change more in the direction of their essays on a subsequent attitude measure than a group that received a D.

In applying the hypothesis of the reinforcing effects of grades to attitude changes in the present context, the formulation presented by Doob (1947) is being followed. Doob views

an attitude as an implicit, anticipatory response which mediates overt behavior but which in turn is derived from the reinforcement of overt behavior. For Doob, reward or avoidance of punishment may constitute the reinforcing contingency. Accordingly, the reinforcement of attitude related, overt statements may be expected to be functionally related to changes in measured attitude. Evidence supporting this view has been supplied by Scott (1957, 1959a, 1959b). The present study attempts to extend Doob's formulation into the area of educational practices.

Scott has proposed the possibility that verbalization of a position opposed to initial attitude in and of itself may produce change with reinforcement or nonreinforcement leading to stability or extinction of the new response. Janis and King (1954) showed that verbalization alone produced change. The relative contribution of verbalization alone in contrast to the effect of the consequent contingencies could not be examined by Scott because of the absence of a control group which experiences no consequences of their verbalizations. In the present experiment, a group was included that received no grade following their essays.

### METHOD

A 40-item questionnaire containing four 10-item attitude scales were administered to 228 students enrolled in Communication Skills classes at the State University of Iowa. Subjects responded to each item on a five-point continuum ranging from strongly agree to strongly disagree. The four attitude scales dealt with federal aid to education, legalized

<sup>1</sup> Now in the Department of Speech, Sacramento State College.

<sup>2</sup> Now in the Department of Speech, University of Connecticut.

gambling, capital punishment, and socialized medicine. Responses to the scale on federal aid to education were extremely skewed and the capital punishment scale proved to be unreliable. These two scales were discarded from further consideration and served primarily for filler items. The initial administration of the questionnaire was conducted by the class instructors.

During a class period, approximately 6 weeks following administration of the scales, the subjects were asked by the experimenter to write essays on particular assigned topics. The subject was instructed by directions appearing at the top of a sheet to write a brief essay supporting a position on either legalized gambling or socialized medicine. Scores on the attitude scales determined the position that was assigned. In each case, the subject was instructed to write supporting the position opposite to that indicated by his pretest scale score, i.e., if the subject's scale score indicated favorability to legalized gambling, he was asked to write in opposition to legalized gambling. The designation of topic to a particular subject was based on the strength of his initial position on the scales. The scale chosen was the one on which the subject had assumed the strongest position relative to the other scales. One-half hour was permitted for the essay. On completion of the essays, the experimenter promised to return grades on the following day.

On a random basis, grades were assigned to the essays. One third of the subject's writing on each topic received a grade of A, one third received a grade of D, and one third was given no grade. This last group was told when papers were returned that the ratings were not completed due to insufficient time. Immediately after returning the essays and grades, the attitude questionnaire was re-administered. Finally, subjects were asked to indicate their satisfaction with the essays. The total number of subjects participating in all phases of the study numbered 127, of whom 58 wrote an essay on legalized gambling and 69 wrote on socialized medicine.

Scores on each 10-item attitude scale were computed by summing the responses to each item for which a weight from 1 to 5 was given. The total minimum possible score for each scale was 10 and the maximum possible score was 50. The dependent change measure was derived by subtracting each subject's score on the posttest from his score on the pretest. A change in the direction of the position taken on the essay was given a positive sign. Change in the opposite di-

rection was negative. To eliminate negative numbers a constant of 20 was added to all change scores.

## RESULTS

It was predicted that subjects who were awarded an A would change on the average in the direction of their essays to a greater extent than subjects who were given a D. Table 1 presents the data relevant to this prediction. Subjects receiving an A changed an average of 31.76 points in the direction of their essays while subjects who were given D changed 25.85 points. This difference is significant at beyond the .01 level. Comparisons of the groups that received a grade with the group that did not, indicates significantly greater change ( $p < .05$ ) for the subjects who received an A than for those given no grade, while no difference is suggested between the subjects obtaining a D and subjects receiving no grade.

Analysis of mean change for each of the issues suggests similar results to those obtained in the overall analysis although the differences obtained were more striking for the socialized medicine issue than for legalized gambling. An analysis of mean change in relation to initial position indicates that those

TABLE 1  
MEAN ATTITUDE CHANGE

Change <sup>a</sup>	Grade A (N = 42)	Grade D (N = 47)	No Grade (N = 38)
Mean	31.76	25.85	27.11
SD	11.16	7.84	8.03
Differences between groups <sup>b</sup>			
A vs. D:	2.84	<.01	
A vs. No:	2.13	<.05	
D vs. No:	0.73	>.10	

<sup>a</sup> A constant of 20 was added to all change scores.

<sup>b</sup> Because of heterogeneity of variance, *t* tests were computed employing the procedure recommended by Edwards (1960).

who had initially assumed a favorable position on each of the issues (i.e., in favor of legalized gambling or socialized medicine) changed significantly more ( $t = 4.73, p < .01$ ) than those who were unfavorable. Nevertheless, the effect of grades on attitude change remains similar to that obtained when direction of initial position is not considered.

Subjects, in addition to responding to the attitude scale related to the topic on which they had written an essay, also responded to the scale for which a related essay had not been written. By comparing the change scores of subjects who had written an essay on a particular topic with those who had not written on that topic it is possible to evaluate the effect of essay writing independent of the effect of grades.<sup>3</sup> The mean change obtained for subjects who had written a relevant essay was 28.24 in contrast to 22.61 for subjects who had written on the other topic. This difference is significant ( $t = 3.27, p < .01$ ) suggesting that the writing of an essay, independent of grade received, produced change in attitude.

#### DISCUSSION

The results suggest support for the hypothesis that a "good" grade serves to reinforce the behavior for which it has been administered. Verbalization without a consequent contingency seems to lead to responses similar to those obtained when verbalization is followed by a "bad" grade. "Cognitive contact" with the opposing side in and of itself does not appear to produce

<sup>3</sup> Comparison of the scores of each of the experimental groups individually with this nonessay writing "control" group would seem to be precluded because of potential unknown effects of simply receiving a grade. Collapsing across groups regardless of experimental condition potentially randomizes such effects.

effects approaching those obtained when reward is forthcoming.

Some insight into additional effects of the grades is provided by the results associated with a question as to the satisfaction the subjects experienced with their essays. Of the 52 subjects who received an A, 42 indicated satisfaction with their essays. Twelve of the 57 subjects who received a D were satisfied with their essays. Of the subjects who received no grade, 12 out of 57 were satisfied. A chi square of 41.99 ( $df = 2$ ) indicated that this distribution is significant beyond the .001 level. It is apparent that no difference in satisfaction with their essays appears for those who received a poor grade and those who were given no grade. Both events appear in this instance to be functionally equivalent in their effects in contrast to the effect of a good grade. The no grade condition was an unusual one in class practice and seemed to operate for the subjects as a poor grade, perhaps due to the frustration of failure for the expectancy of receiving some grade to be fulfilled.

Some evidence is provided that verbalization in an incongruent essay is effective in producing attitude change, independent of consequences. This supports the findings of Janis and King (1954) on the relation of role playing and attitude change.

The experiment presents evidence only on the effect of grades on one aspect of a complex set of behaviors appearing in the essay writing situation. A qualified generalization can be offered that the administered grades affected in a similar manner many other unmeasured aspects of performance in the situation, i.e., compositional skills, ideational patterns, affect concerning essay writing, etc. Further research is necessary to delineate those behaviors affected in grading situa-

tions. Educational practice will profit from understanding of the functional role of grades for other than description of assessment procedures.

### SUMMARY

The potential effects of grades as a reinforcing contingency were examined. University students wrote essays defending positions on attitude related issues contrary to their previously assessed positions. Good and poor grades were randomly assigned to the essays and reported to the students. The effect of these procedures on attitude change were evaluated and good grades were demonstrated to serve a reinforcing role in contrast to the effects of a poor grade or no grade.

### REFERENCES

- DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135-156.
- EDWARDS, A. L. *Experimental design in psychological research*. (Rev. ed.) New York: Rinehart, 1960.
- JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211-218.
- ODELL, C. W. Marks and marking systems. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. (Rev. ed.) New York: Macmillan, 1950. Pp. 711-717.
- SCOTT, W. A. Attitude change through reward of verbal behavior. *J. abnorm. soc. Psychol.*, 1957, **55**, 72-75.
- SCOTT, W. A. Attitude change by response reinforcement: Replication and extension. *Sociometry*, 1959, **22**, 328-335. (a)
- SCOTT, W. A. Cognitive consistency, response reinforcement, and attitude change. *Sociometry*, 1959, **22**, 219-229. (b)

(Received October 13, 1960)

# **GRADUATE EDUCATION IN PSYCHOLOGY**

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:

Anne Roe, *Chairman,*

and

John W. Gustad, Bruce V. Moore,  
Sherman Ross, and Marie Skodak

Price \$1.50

**THE AMERICAN  
PSYCHOLOGICAL ASSOCIATION  
Dept. Grad  
1333 Sixteenth Street, N.W.  
Washington 6, D.C.**

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

June 1961

Number 3

## EFFECTS OF EXTENDED EFFORT IN CREATIVE PROBLEM SOLVING<sup>1</sup>

SIDNEY J. PARNES<sup>2</sup>

*University of Buffalo*

Several studies have demonstrated the possibility of the deliberate development of individual creativity through educational programs (Parnes, 1958, 1960; Taylor, 1959). Two experiments evaluated the creative problem solving course (Parnes, 1959) conducted at the University of Buffalo since 1949 (Meadow & Parnes, 1959; Parnes & Meadow, 1960). In the creative problem solving course a variety of principles and procedures are used to develop creative ability. Two studies have demonstrated the effectiveness of the particular brainstorming principle (Osborn, 1957) of deferred judgment during idea production (Meadow, Parnes, & Reese, 1959; Parnes & Meadow, 1959). Taylor studied the use of the principle among groups vs. individuals (Taylor, Berry, & Block, 1957).

An additional result of the Parnes and Meadow (1959) study indicated a positive correlation between quantity and quality of ideas. The relationship between quantity and quality of ideas has been of great interest in the study

of creative problem solving activities. One principle stressed in the course is that copious idea production leads to more creativity in problem solving. Osborn and others have emphasized this need for fluency of ideas. Instructors have reported a prevalence of better ideas at the end of an idea production period than at the beginning (Parnes, 1959). Previous experimenters found increases in remoteness of response and uncommonness of response among succeeding responses on creative tasks (Christensen, Guilford, & Wilson, 1957).

The present study involved two experiments designed to test the principle that extended effort in idea production will lead to an increasing proportion of good ideas with increased production.

### EXPERIMENT I

The hypothesis was formulated in Experiment I that more good ideas will appear in the last half of a subject's total idea output regarding a creative thinking problem than during the first half. It was also decided to test for any difference between high and low producers in the proportions of their good ideas that fall in the second half.

<sup>1</sup>This research was financed by a grant from the Creative Education Foundation.

<sup>2</sup>The author is indebted to Hayne Reese for his constructive review of this paper, as well as for his assistance with the statistical analyses.

## Method

**Subjects.** Data were available from two groups of subjects. Group 1 consisted of 68 students registered (but uninstructed) in the creative problem solving course at the University of Buffalo. Group 2 was comprised of 78 students enrolled in other undergraduate courses at the university. The data for the two groups were not combined because a cursory examination gave reason to suspect that the groups might differ. (Their averages on the experimental problem were 7.9 and 6.7, respectively.) Accordingly, a statistical treatment was applied which allowed the use of all the data, but took into account this possible difference.

**Experimental problem.** The Hanger problem used in the present experiment was selected from Part V of the AC Test of Creative Ability. This test is reported to have differentiated groups of creative from noncreative subjects (Harris & Simberg, 1954). Data are available from a previous experiment which indicated that scores on the Hanger problem are positively correlated with other tests which have been designed to measure creative thinking (Meadow & Parnes, 1959).<sup>3</sup> Instructions for the Hanger problem required subjects to list all possible uses for an ordinary wire coat hanger. Time limit for the test was 5 minutes.

**Ratings.** The test was group administered during regular class periods. The scoring procedure was modified to yield a quantity and uniqueness score instead of a quantity and uniqueness score, as reported in the previous experiment (Meadow & Parnes, 1959). Each response was scored, by a trained research assistant, as indicating either good or bad quality. The quality score was defined as comprising two dimensions: uniqueness—degree to which the response departed from the hanger's conventional use and value—the degree to which the response was judged to have social, economic, esthetic, or other usefulness.

The scorer was instructed to rate each response on a three-point scale for uniqueness and a three-point scale for value. The response was finally scored as indicating good

quality if assigned a combined uniqueness and value score of at least 5. Final quality score used was the total number of "good quality" responses. Any response which duplicated (in essential meaning) responses already given was eliminated from the scoring.

The rater had no knowledge of the purpose of this study; therefore, there was no reason to suspect any bias in ratings of early ideas vs. later ideas. Interrater reliability for ratings of 50 subjects selected at random was .74.

**Procedure.** The subjects were divided into subgroups on the basis of their total quantities of ideas. Group 1 was divided at its median of 7.9, and Group 2 at its median of 6.7. The subjects with a total number of ideas less than the median were assigned to Subgroup L, and those with a total number of ideas greater than the median were assigned to Subgroup H. To obtain proportional cell frequencies for the analysis of variance, the data of 12 subjects, 6 from each group, were eliminated in a random manner, with the restriction that their total number of ideas was near the median. There remained 62 Group 1 and 72 Group 2 subjects.

Each subject's data were divided into halves on the basis of his total number of ideas. Those subjects with an odd number of ideas were treated as follows: If the middle idea was a good one, it was divided between the halves by adding one-half point to the score for each half.

## Results and Discussion

Table 1 presents the mean number of "good" ideas in each half for each group. There were more good ideas in the second half than in the first half, and more good ideas in Subgroup H than in Subgroup L, but there was little difference between Group 1 and Group 2. A Lindquist Type III analysis of variance of these data, (Lindquist, 1956) indicated highly significant differences between Half I and Half II ( $F = 32.16$ ;  $df = 1,130$ ;  $p < .001$ ) and between Subgroups L and H ( $F = 56.37$ ;  $df = 1,130$ ;  $p < .001$ ). No other differences were statistically reliable.

The results of the experiment indicate that extended effort in producing ideas on a creative thinking problem tends to reward subjects with a greater

<sup>3</sup> Scores on the Hanger problem in the previous experiment correlated with other creative ability tests as follows: Guilford's Unusual Uses, .473; Guilford's Plot Titles High, .452; Guilford's Apparatus, .301; TAT Originality, .520. All but the Apparatus correlation were at the .01 level of significance. The correlation with the Apparatus test was significant at the .05 level.

proportion of good ideas in the second half of their total output. As in a previous experiment (Parnes & Meadow, 1959) a significant relationship was found between quantity and quality of ideas. There was no significant difference between high and low producers in their proportions of good ideas in the second half of total production.

The findings support Osborn's theory (1957) that in idea production, quantity leads to quality. The results also seem to be in accordance with Gordon's explanation of "deferment" in the creative process. He describes deferment as "the capacity to discard the glittering immediate in favor of a shadowy but possibly richer future" (Gordon, 1956). The *noncreative* problem solver gets an idea, sees it as a possible solution to his problem, and settles for it without further ado. The *creative* problem solver is not satisfied with his first idea. Like the person who invests money to obtain greater rewards later, the creative person forgoes the immediate reward of applying his first idea, in expectation of a better solution (greater reward) ultimately. A further hypothesis suggested by Osborn's and Gordon's theories is that the *best* idea will come late in the total production period. The author is currently considering experiments to test this particular hypothesis.

## EXPERIMENT II

Experiment II was designed to determine whether the type of results found in Experiment I with untrained subjects would also occur with subjects trained in the use of the brainstorming principle of deferred judgment. It was also decided to lengthen the time period to 15 minutes instead of 5, in order to see if a trend could be observed towards increasingly greater

TABLE 1  
MEAN NUMBER OF GOOD IDEAS AMONG  
FIRST AND SECOND HALVES OF ALL  
IDEAS AND AMONG ALL IDEAS

Group	First Half	Second Half	All Ideas
Group 1			
Low	.69	1.11	1.81
High	1.61	2.48	4.10
Mean	1.15	1.80	2.95
Group 2			
Low	.51	1.24	1.75
High	1.47	2.42	3.89
Mean	.99	1.83	2.82
Combined			
Low	.60	1.18	1.78
High	1.54	2.45	3.99
All Subjects	1.07	1.81	2.88
Mean			

Note.—Low refers to the group whose total number of ideas was less than the median number of ideas; High refers to the group whose total number of ideas was greater than the median number of ideas.

proportions of good ideas as a subject's total quantity increases.

### Method

*Subjects.* Forty-two students in two creative problem solving classes at the University of Buffalo served as subjects. The experiment was conducted during the final half of the course.

*Experimental problem.* The Hanger problem was used as in Experiment I. However, the trained students were very familiar with the brainstorming principle. Therefore, they were urged to follow this principle of deferred judgment during their idea production. Time limit was 15 minutes.

*Ratings.* The same rating system was used as in Experiment I. However, before the new experiment was conducted, the rater had already been told the purpose of the original study. If he were given the subjects' protocols, his ratings of the ideas might have been influenced this time by the relative position of each idea on the page. Accordingly, the rater was given each subject's list of ideas typed onto new sheets in alphabetical order. Thus he did not know the position of an idea in a subject's total list of ideas. After the rater had scored the ideas on the

typed sheets, clericals transferred these ratings back to the subjects' original lists.

*Procedure.* The total list of ideas of each subject was then divided into thirds—a first third, a middle third, and a last third in a similar manner as was used to divide the total ideas into halves in Experiment I.

### Results and Discussion

The mean numbers of good ideas per third were 4.84, 5.09, and 5.81 for the first, middle, and last thirds, respectively. The analysis of variance of these data indicated that the main effect of thirds was significant ( $F = 4.12$ ;  $df = 2,84$ ;  $p < .05$ ). *t* tests showed that the mean for the last third was significantly greater than the means for both the first third ( $t = 2.75$ ;  $df = 84$ ;  $p < .01$ ) and the middle third ( $t = 2.04$ ;  $df = 84$ ;  $p < .05$ ). The means for the first and middle thirds did not differ significantly ( $t < 1.00$ ). Thus the findings of Experiment I are supported, and a trend is suggested towards increasingly larger proportions of good ideas with increased quantity. The results seem to provide further interest in the hypothesis that the *best* idea will come late in the total production period.

The data of the first 5 minutes only<sup>4</sup> were examined on the same basis as Experiment I. For these first 5 minutes, a mean of 3.02 good ideas were produced in the first half of total ideas and a mean of 3.26 in the second half. Analysis of variance of these data indicated that these means were not significantly different. It is interesting to note that these means are both almost double to triple the means for the untrained subjects reported in the results section of Experiment I. This supports the evidence of the earlier experiments which demonstrated the effectiveness of the brainstorming principle and of

training in its use (Meadow et al., 1959; Parnes & Meadow, 1959).

### DISCUSSION

Examination of the protocols suggested that the untrained subjects of Experiment I seem to start their production with more of the conventional, unoriginal ideas—high in usefulness but low in uniqueness. Toward the end of a 5-minute period, however, these untrained subjects may find themselves exhausted of these ideas, and may begin groping for less obvious ones. This may lead to their greater production, among the later ideas, of ones high in both usefulness and uniqueness, the criteria for the "good" ideas. This seems to be consistent with the findings of Christensen et al. (1957), where uncommonness and remoteness of responses increased with succeeding responses. In the protocols of the trained subjects of Experiment II, there seem to be many of the unique and useful ideas among the first ones; but within a 15-minute period these trained subjects may also tend to exhaust themselves of ready ideas and find that they too have to "stretch" their imaginations more fully. This may explain why their strongest production of good ideas in the 15-minute period does not appear until the last third of their ideas.

In order further to explore the difference in results between trained and untrained subjects, a sign test analysis was made of scored data on hand from 180 untrained subjects of a previous experiment (Meadow & Parnes, 1959), for an entirely different type of problem (Guilford's Plot Titles) with a 3-minute test period. These data indicated no significant difference in number of quality ideas produced in first half (mean of .57) vs. second half (mean of .62) of the idea lists.

Thus we have data suggesting: (a)

<sup>4</sup> Subjects had been requested at periodic time intervals during Experiment II to draw a line under their last idea.

for *untrained* subjects, significantly more good ideas among *later* ideas in a 5-minute period, but not significantly more in a 3-minute stint; (b) for *trained* subjects, significantly more good ideas among *later* ideas for a 15-minute period, but not significantly more for a 5-minute stint. Theorizing from the four sets of related data, it might be suggested that subjects will get at least as many good ideas in the second half of their total idea output regarding a creative thinking problem as in the first half, but that with increasing production time, the proportion of good ideas in the second half will increase over the first half. It might be further theorized that untrained subjects will show this increasing proportion more rapidly than trained subjects, who have already learned to be more creative in their early ideas. In a sense, both trained and untrained subjects seem to be stimulating their creativity by *extended* effort in idea-finding. For the trained subjects, a 5-minute period (on the type of problem used) does not seem to be long enough to elicit *extended* effort on their parts. (This might be related analogously to the accomplished musician practicing on a beginner's piece.) However, for untrained subjects 5 minutes *did* seem to provide enough time to call for *extended* effort. Much further experimentation is needed with both trained and untrained subjects, given a variety of time limits and types of problems, in order to verify the theory suggested.

### SUMMARY

Experiment I was designed to test the hypothesis that more good ideas will appear in the last half of a subject's total idea output regarding a problem than during the first half. Each subject was given a creative thinking problem which involved pro-

duction of ideas for a 5-minute period. The resultant ideas were evaluated by a trained rater. Number of good ideas was tallied for the first half of each subject's total idea list, and for the second half of his list.

The findings demonstrated significantly more good ideas to appear in the last half than in the first half of the idea lists. A significant relationship was also found between total quantity and total quality scores.

Experiment II was designed to determine whether the type of results found in Experiment I with untrained subjects would also occur with subjects trained in the use of the brainstorming principle of deferred judgment. It was also decided to lengthen the time period to 15 minutes instead of 5, in order to see if a trend could be observed towards *increasingly* greater proportions of good ideas as a subject's total quantity increases. Each subject was given 15 minutes to produce ideas regarding the same creative thinking problem as in Experiment I. Resultant ideas were evaluated as in the first experiment. Number of good ideas was tallied for each third of each subject's total idea list—first third, middle third, and last third.

The findings showed significantly more good ideas to appear in the final third of the subjects' idea lists. Although there was not a significant difference between first and second "thirds," a trend is suggested towards increasingly larger proportions of good ideas with increased quantity.

The results are interpreted to indicate that extended effort in producing ideas on a creative thinking problem tends to reward problem solvers with a greater proportion of good ideas among the later ideas on their lists.

### REFERENCES

- CHRISTENSEN, P. R., GUILFORD, J. P., & WILSON, R. C. Relations of creative re-

- sponses to working time and instructions. *J. exp. Psychol.*, 1957, **53**, 82-88.
- GORDON, W. J. J. Operational approach to creativity. *Harv. bus. Rev.*, 1956, Nov.-Dec., 41-51.
- HARRIS, R. R., & SIMBERG, A. L. *AC Test of Creative Ability: Examiner's manual*. Detroit: General Motors Corp., AC Spark Plug Division, 1954.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
- MEADOW, A., & PARNE, S. J. Evaluation of training in creative problem solving. *J. appl. Psychol.*, 1959, **43**, 189-194.
- MEADOW, A., PARNE, S. J., & REESE, H. Influence of instructions and problem sequence on a creative problem solving test. *J. appl. Psychol.*, 1959, **43**, 413-416.
- OSBORN, A. F. *Applied imagination*. New York: Scribner, 1957.
- PARNE, S. J. (Ed.) *Compendium of research on creative imagination*. Buffalo, N.Y.: Creative Education Foundation, 1958.
- PARNE, S. J. *Instructor's manual for semester courses in creative problem-solving*.
- Buffalo, N.Y.: Creative Education Foundation, 1959.
- PARNE, S. J. (Ed.) *Compendium No. 2 of research on creative imagination*. Buffalo, N.Y.: Creative Education Foundation, 1960.
- PARNE, S. J., & MEADOW, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *J. educ. Psychol.*, 1959, **50**, 171-176.
- PARNE, S. J., & MEADOW, A. Evaluation of persistence of effects produced by a creative problem-solving course. *Psychol. Rep.*, 1960, **7**, 357-361.
- TAYLOR, C. W. (Ed.) *The third University of Utah research conference on the identification of creative scientific talent*. Salt Lake City, Utah: Univer. Utah Press, 1959.
- TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking. *Admin. sci. Quart.*, 1958, **3**, 23-47.

(Received October 3, 1960)

## PATTERNS OF SELF-CONCEPTUALIZATION IN HIGH SCHOOL AND COLLEGE STUDENTS

WILBUR C. WASHBURN

*University of California*

Self-concepts have been related to aspects of child development in the theoretical orientations of various psychologists—Freud, Erikson, Sarbin, Horney, and Fromm. Although the specific term "self-concept" has not always been used, the "character traits" described by these writers correspond to what is now frequently called a self-concept. For purposes of this paper, a self-concept is regarded as a relatively enduring aspect of personality—a motive, attitude, or value—by means of which the individual relates himself to his social environment. The term applies to the person's sets of ideas about himself in relation to other persons.

### PROCEDURE

The problem was to find patterns of self-concepts useful in describing adolescents and young adults and to integrate the various theories that formulate self-concepts in terms of child development. Since many existing sets of postulated self-concepts were originally inferred on the basis of very early developmental periods, they may not be adequately descriptive of more mature levels of personality. The procedure, then, was to translate the hypothetical patterns of self-concepts offered by a variety of theorists into test items, to administer the test to groups of adolescent and young adult subjects, and to revise the list of hypothetical patterns of self-conceptualization on the basis of intercorrelation of test items. The resulting sets of self-concepts were then ranked on the basis of maturity ratings provided by

groups of graduate students specializing in developmental psychology.

Since the results of any statistical analysis of test items depend in part on the nature of the items selected for analysis, an effort to minimize this limitation was made by including items based on a variety of theoretical viewpoints. Table 1 indicates the postulated patterns of self-concepts for which items were devised and lists a sample item for each category. Items were based on the following theoretical formulations: (a) the empirical selves hypothesized by Sarbin (1952), (b) the psycho-social stages stated by Erikson (1960), (c) the character types described by Freud (1933), (d) the personality patterns postulated by Fromm (1947), and (e) the interpersonal attitudes developed by Horney (1945).

### *Construction of the Testing Instrument*

The test originated with a cluster analysis, using the technique described by Tryon (1939), applied to Sarbin's five empirical selves. The author constructed 135 items representing self-concepts postulated by Sarbin. Five graduate students specializing in personality theory were asked to classify each item into the category it was supposed to measure. All five judges correctly classified 110 items, which then were administered to a group of 100 high school students.

Four patterns of self-concepts emerged as a result of the cluster analysis. Sarbin's first three empirical selves were highly intercorrelated and

TABLE 1

## THE POSTULATED PATTERNS OF SELF-CONCEPTS ON WHICH TEST ITEMS WERE BASED

Personality Theorist	Hypothesized Self-Concept	Example of Item <sup>a</sup>
Erikson	Infancy (Trust vs. Mistrust) Early Childhood (Autonomy vs. Shame and Doubt) Play Age (Initiative vs. Guilt) School Age (Industry vs. Inferiority) Adolescence (Identity vs. Identity Diffusion) Young Adulthood (Intimacy vs. Isolation) Adulthood (Generativity vs. Self-Absorption) Senescence (Integrity vs. Disgust)	Most persons can be trusted. I prefer to do things for myself.  I take the lead in getting friends together. I really enjoy working.  I usually stick to my own ideas even when friends disagree. When feeling blue I try to find someone who is cheerful. I wish I could discover something that would benefit future generations. I try to find my faults and correct them.
Freud	Oral Character Anal Character Genital Character	I look on the bright side of things. I tend to save my money. I love humanity in general.
Fromm	Receptive Character Exploitative Character Hoarding Character  Marketing Character Productive Character	I prefer letting others make plans. I want to win when pitted against others. If buying a car, I would be most interested in its economy. It is important to always please others. Friends say I am creative and original.
Horney	Compliant (Moving Toward) Detached (Moving Away) Aggressive (Moving Against)	I enjoy just being with people. I prefer being alone. I often object to another's ideas.
Sarbin	Somatic Self Receptor-Effector Self Primitive-Construed Self Introjecting-Extrojecting Self  Social Self	My body has greatly influenced me. I can't wait to eat when hungry. I let off steam when angry. I usually do the expected thing to avoid disapproval of others. If I cheated on a test, I would feel so guilty I couldn't stand it.

<sup>a</sup> Items were selected randomly for illustration.

were labeled the Somatic-Primitive Self. Items from Sarbin's fifth empirical self had a high negative correlation with the first cluster and were called the Inner-Controlled Self. Sarbin's fourth empirical self split into two sets of items. One cluster was labeled the Submissive-Dependent Self, while the other was tentatively named the Socially Integrated Self. These labels were assigned each cluster by considering apparent logical relationships among items.

Items representing a wider variety of theories were then added. The au-

thor constructed 250 items based on the theories of Erikson, Freud, Fromm, and Horney. The judges were again asked to classify each new item into the category or categories it presumably measured, and the 225 items on which there was 100% agreement were administered to a second group of 100 high school students. A biserial coefficient of correlation was computed to show the relationship of each item to the total score of the four existing clusters. Items were assigned to the cluster with which they had the highest correlation.

All items representing Horney's compliant orientation and many items describing Fromm's receptive and marketing characters were added to the test measuring the Submissive-Dependent Self. A large number of items were added to the test measuring the Inner-Controlled Self from those representing Freud's anal character, Fromm's hoarding type, and the problems listed by Erikson under school age, adolescence, adulthood, and senescence. The cluster labeled Socially Integrated Self absorbed many items representing Freud's oral and genital characters, Fromm's productive character, and the problems stated by Erikson under adulthood and senescence. Inspection of the items added to this cluster suggested a similarity to the self-actualizing personality described by Maslow (1954), and the cluster was renamed the Integrative-Actualizing Self. A set of items was found to have a high negative correlation with the Submissive-Dependent Self and was labeled the Outer-Controlling Self. This new pattern of self-concepts was composed largely of items representing Horney's aggressive orientation, although it also contained items measuring the problems described by Erikson as typical of early childhood, the play age and the school age. Another group of items had a high negative correlation with the Integrative-Actualizing Self and was

named the Detached-Independent Self. It was composed of items measuring Horney's detached orientation and the problems outlined by Erikson with regard to the periods of infancy, early childhood, and young adulthood.

Table 2 indicates relatively high negative correlations between certain of the six resulting patterns of self-concepts. Each pair of clusters with high negative correlations described contrasting self-concepts. Hence forced-choice items were developed in which items from negatively correlated clusters were combined as alternative choices in a single question. Subtest 1 measured the Somatic-Primitive Self and the Inner-Controlled Self. Subtest 2 measured the Submissive-Dependent Self and the Outer-Controlling Self. Subtest 3 measured the Detached-Independent Self and the Integrative-Actualizing Self. Sample forced-choice items from each subtest are presented in Table 3.

The use of forced-choice items offered several advantages. It resulted in a 50% reduction in the number of answers to be scored on each answer sheet. Once a score was obtained for a given cluster, the score for its negatively correlated counterpart could be computed simply by subtracting the first score from the maximum attainable score. Anastasi (1954) points out that this testing technique provides changing alternatives and thus avoids

TABLE 2  
INTERCORRELATIONS AMONG SELF-PATTERNS

	Submissive-Dependent Self	Detached-Independent Self	Outer-Controlling Self	Inner-Controlled Self	Integrative-Actualizing Self
Somatic-Primitive Self	-.17	.10	.21	-.83	-.12
Submissive-Dependent Self		.10	-.86	.18	-.08
Detached-Independent Self			-.09	-.01	-.78
Outer-Controlling Self				-.19	.11
Inner-Controlled Self					.13
Integrative-Actualizing Self					

TABLE 3  
SAMPLE ITEMS FROM THE THREE SUBTESTS

Subtest	Examples of Test Items <sup>a</sup>	
1		
(A) Inner-Controlled Self	(A) If I won \$1,000, I would be more likely to save most of it for future needs.	(B) If I won \$1,000, I would be more likely to spend most of it now on things I need to have a good time.
(B) Somatic-Primitive Self	(A) I would prefer having parents, teachers, or employers think well of me.  (A) My family has had a greater influence on me.  (A) I say little when angry.  (A) I am sure most policemen are helpful.	(B) I would prefer just taking it easy and enjoying life.  (B) My body and its feelings have had a greater influence on me.  (B) I let off steam when angry.  (B) It is likely too many policemen try to show their authority.
2		
(A) Outer-Controlling Self	(A) I like giving directions to others.	(B) I prefer following directions given by others.
(B) Submissive-Dependent Self	(A) I want to be considered strong by others.  (A) I would like to be thought of as an important person in my community.  (A) I often raise objections to another's ideas.	(B) I prefer being thought of as kind and sympathetic.  (B) I wouldn't care to be considered an important person.  (B) I usually accept the ideas of others without being critical.
3		
(A) Integrative-Actualizing Self	(A) My former teachers would be more likely to call me a leader.	(B) My former teachers would be more likely to call me a follower.
(B) Detached-Independent Self	(A) I often enjoy change and rather look forward to trying new things and meeting new people.  (A) One of my best features is that I like people.  (A) I realize human nature isn't perfect, but I can accept it and even like it as it is.  (A) I am left with a feeling of strength upon seeing or hearing something truly beautiful.  (A) I am usually optimistic and enthusiastic and seldom lonely and depressed.	(B) I dislike change and tend to feel uncomfortable in new situations.  (B) One of my best features is that I don't let feelings influence my judgment.  (B) I find it hard to accept human nature as it is.  (B) I am not too impressed by art or music, although these things seem beautiful to others.  (B) There are quite a few times when I feel alone and afraid and just about the whole world seems unfriendly.

<sup>a</sup> Items were selected randomly for illustration.

the tendency toward a response set characteristic of constant-alternative, true-false items.

Forced-choice alternatives were selected from items which seemed to logically fit together. However, items chosen by over 90% or fewer than 10% of the total group were eliminated, since they seemed to reflect extremes of social approval and disapproval. New forced-choice items, 50 for each subtest, were added to existing items and the revised test administered to a third group of 100 high school students. A biserial coefficient of correlation was computed showing the relationship of each item to the total score. Items correlating less than .20 with the total subtest were eliminated. The new forced-choice form was then administered to a group of 300 undergraduate students at the University of California, each item correlated with the total score, and additional items eliminated.

Validity of the subtests was appraised by asking high school counselors to rate students for each subtest on the basis of an adjective checklist describing characteristics appropriate for the subtest. For example, descriptive terms applicable to the first subtest included the following words: conscientious, controlled, conforming. Subjects with ratings in the upper and lower 27% of the group based on total number of adjectives checked for each subtest were then compared for mean scores on the subtest and *t* ratios computed. The differences between means were all significant at less than the .01 level of confidence. Similar results were obtained in a second validation study. Corrected split-half reliability coefficients were above .80 for groups from the tenth grade to college undergraduates.<sup>1</sup>

Five graduate students at the University of California were asked to rank each cluster in terms of relative maturity as inferred on the basis of their knowledge of developmental psychology. Each judge had completed at least two advanced undergraduate courses and graduate research in this area. A second group of five graduate students with similar qualifications but from a different institution also ranked the six patterns of self-concepts. While both groups agreed on the rank order, agreement was less consistent for the relative position of the Detached-Independent Self and the Outer-Controlling Self than for the other levels of self-conceptualization. There was 100% agreement on the Somatic-Primitive Self; 90% each on the Dependent-Submissive Self, the Inner-Controlled Self, and the Integrative-Actualizing Self; 70% on the Detached-Independent Self; and 60% on the Outer-Controlling Self.

The six resulting self-patterns, listed from the one rated least to the one ranked most mature, may be described as follows: (a) the Somatic-Primitive Self involves inability to postpone need satisfaction and failure to identify with authority figures, (b) the Submissive-Dependent Self involves avoidance of disapproval by submitting to others, (c) the Detached-Independent Self involves avoidance of interpersonal relations and situations likely to arouse emotion, (d) the Outer-Controlling Self involves achievement of social status through manipulation of the external world in socially approved ways, (e) the Inner-

mentation Institute. Order Document No. 6612 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>1</sup> A copy of the testing instrument has been deposited with the American Docu-

Controlled Self involves development of inner controls through internalization of external social norms, (f) the Integrative-Actualizing Self involves acceptance of oneself and others and a tendency toward increased creativity and productivity.

### RESULTS

#### *A Comparison of High School and College Groups*

The final form of the test was utilized with a sample of high school and college students. The high school students were from a large urban high school in California and were randomly selected within categories to constitute a stratified sample proportionate to the college group regarding racial and socioeconomic classification. The college students consisted of undergraduates attending the University of California at Berkeley, and the sample was selected to be representative of this campus with regard to fields of study. Separate results are

reported for males and females in both groups and for college preparatory and noncollege preparatory high school students. The college preparatory group consisted of students taking courses which will enable them to attend a private or state college, a private or state university, or a junior college. The college sample constitutes a group selected from a sample similar to the college preparatory group on the basis of academic achievement in high school. Further details regarding these groups are presented in Table 4.

Means for high school and college groups are indicated in Table 5. Separate means are not given for each age group within the high school or college groups, since no statistically significant differences were found with regard to such age levels. Since each subtest measures two negatively correlated self-levels, one rated as relatively more mature than the other, the means presented in Table 5 are for the self-pattern with the higher maturity ranking. For example, the

TABLE 4  
COMPOSITION OF GROUPS USED IN THE STUDY

Characteristics	College Sample	High School: Noncollege Preparatory	High School: College Preparatory
Males	66	40	100
Females	49	45	135
Median Age (Males)	22 yrs.-1 mo.	16 yrs.-11 mo.	16 yrs.-10 mo.
Median Age (Females)	21 yrs.-10 mo.	16 yrs.-11 mo.	16 yrs.-9 mo.
Caucasian	99	86%	202
Negro	9	8%	19
Other	7	6%	14
Upper Socioeconomic Status <sup>a</sup>	32	28%	66
Middle Socioeconomic Status <sup>a</sup>	60	52%	122
Lower Socioeconomic Status <sup>a</sup>	23	20%	47
First Year	15	0	0
Second Year	61	0	0
Third Year	20	30	92
Fourth Year	19	55	143

<sup>a</sup> Based on father's occupation.

TABLE 5  
A COMPARISON OF HIGH SCHOOL AND COLLEGE GROUPS

	N	Subtest 1		Subtest 2		Subtest 3	
		M	SD	M	SD	M	SD
<b>Males</b>							
High School, Noncollege	40	70.1	16.3	50.4	11.6	70.0	15.1
High School, College Preparatory	100	78.2	25.8	58.4	13.9	74.2	20.7
College	66	79.3	16.2	66.7	11.9	76.3	16.6
<i>t</i> (High School, Noncollege vs. High School, College Preparatory)			1.81		3.18**		1.15
<i>t</i> (High School, College Preparatory vs. College)			.25		3.44**		.57
<b>Females</b>							
High School, Noncollege	45	71.5	20.4	41.0	15.1	77.5	14.3
High School, College Preparatory	135	79.0	19.1	50.1	14.5	86.6	16.0
College	49	78.0	16.7	55.9	14.4	71.5	19.2
<i>t</i> (High School, Noncollege vs. High School, College Preparatory)			2.46		3.65**		2.27
<i>t</i> (High School, College Preparatory vs. College)			.62		2.60		4.20**
<b>Males-Females Compared</b>							
<i>t</i> (High School, Noncollege Females vs. High School, Noncollege Males)			.33		3.06**		2.35
<i>t</i> (High School, College Preparatory Females vs. High School, College Preparatory Males)			.58		4.39**		3.91**
<i>t</i> (College Females vs. College Males)			.31		4.38**		1.42

\*\* Significant at the .01 level of confidence or less.

means given under Subtest 1 are for the Inner-Controlled Self, and the means for the Somatic-Primitive Self can be determined by subtracting the stated means from the maximum possible score of 124 points. Means presented under Subtest 2 apply to the Outer-Controlling Self, and means for the Submissive-Dependent Self can be determined by subtracting the stated means from 103, the highest attainable score for this subtest. Means listed for Subtest 3 represent the Integrative-Actualizing Self, and means for the Detached-Independent Self may be

computed by subtracting the stated means from the maximum possible score of 120 points.

Although high school boys preparing for college tended to score higher than those not preparing for college on all three self-levels rated most mature, the difference was significant at less than the .01 level only for the Outer-Controlling Self. Although college males scored higher than high school college preparatory males on all three mature levels of self-conceptualization, differences were slight and not significant at an acceptable level ex-

cept for the Outer-Controlling Self, which involved a difference significant at less than the .01 level.

High school girls preparing for college scored higher than noncollege preparatory females on the three self-levels with highest maturity rankings. Differences were significant at less than the .02 level for the Inner-Controlled Self, at less than the .01 level for the Outer-Controlling Self, and at less than the .05 level for the Integrative-Actualizing Self. Means representing the Outer-Controlling Self were higher for college women than college preparatory high school girls, and the difference was significant at about the .01 level. However, female college students obtained a lower mean than college preparatory high school girls on the test measuring the Inner-Controlled Self, although the difference was not statistically significant at an acceptable level. The college females also had a lower mean than college preparatory females on the test measuring the Integrative-Actualizing Self, and the difference was large and significant at less than the .01 level.

Both groups of high school females scored higher than corresponding groups of males on the test measuring the Inner-Controlled Self, while males scored slightly higher than females on this test at the college level. However, these differences were neither large nor statistically significant. All male groups scored higher than corresponding female groups on the test measuring the Outer-Controlling Self, and all differences were relatively large and significant at less than the .01 level. High school girls had higher means than high school boys for the Integrative-Actualizing Self. The difference between means on this test was significant at almost the .02 level for noncollege preparatory students

and at less than the .01 level for college preparatory students. The trend was reversed at the college level, and college men tended to score higher than college women on this test, although the difference was not statistically significant.

#### DISCUSSION

It is to be expected that college students would tend to score higher than high school students on tests measuring mature self-levels as a result both of college experience and selection of more mature students for admission to the university. This hypothesis tends to be confirmed for males with regard to all three more mature self-levels, although differences are large and significant at less than the .01 level only for the Outer-Controlling Self. Further, females do not seem to follow this trend in relation to the Inner-Controlled Self and the Integrative-Actualizing Self. However, these relatively low scores of University of California females indicate the female group is more highly select than the college males used in this study. Statistics for the high school district show that about 1.5 times more females than males from this area enter private colleges rather than attend public universities. The emphasis on academic achievement as an entrance requirement at the university possibly results in selection of more competitive individuals and would account in part for the relatively high scores of both male and female college groups on the test measuring the Outer-Controlling Self. The tendency for males consistently to score higher than females on the test measuring the Outer-Controlling Self is to be expected in view of child training practices in our society which from early childhood permit males greater freedom than females in exploring and manipulating their en-

vironments. Since females mature both physically and socially at an earlier age than males, it is not surprising to find that high school females tend to score higher than males on the two self-patterns with the highest maturity rankings, while the trend is reversed and males score higher than females at the university level.

The findings for high school groups indicate that college preparatory students tend to have more mature self-concepts than noncollege preparatory students. The large and statistically significant differences obtained between college preparatory and non-college preparatory students on the test measuring the Outer-Controlling Self suggest that college preparatory students show greater drive to attain status long before attending college. If subsequent research confirms differences in self-concepts related to college attendance and preparation for college, it becomes important to consider personality factors along with academic achievement as a basis for predicting future success among college applicants.

Since the results are likely to vary with the theoretical viewpoints represented in the test items analyzed, the self-patterns described in this study should not be regarded as a final list. Further, the six self-levels should not be interpreted as rigid categories into which all persons may be classified on an all-or-nothing basis. Every subject checked at least some items for every one of the six self-patterns. Hence it seems safe to assume that the more immature levels continue into at least

early adulthood to some extent in all individuals.

#### SUMMARY

Items measuring self-concepts were constructed representing the theories of Sarbin, Freud, Erikson, Horney, and Fromm. A cluster analysis yielded six self-patterns, which, ranked from least to most mature on the basis of ratings by graduate students in psychology, were: (a) the Somatic-Primitive Self, (b) the Submissive-Dependent Self, (c) the Detached-Independent Self, (d) the Outer-Controlling Self, (e) the Inner-Controlled Self, and (f) the Integrative-Actualizing Self. Significant differences in means on tests measuring various self-levels were found for contrasting groups of males and females, college preparatory and non-college preparatory high school students, and college and high school students.

#### REFERENCES

- ANASTASI, A. *Psychological testing*. New York: Macmillan, 1954.
- ERIKSON, E. H. Youth and the life cycle. *Children*, 1960, 7(2), 43-49.
- FREUD, S. *New introductory lectures on psychoanalysis*. New York: Norton, 1933.
- FROMM, E. *Man for himself*. New York: Rinehart, 1947.
- HORNEY, K. *Our inner conflicts*. New York: Norton, 1945.
- MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.
- SARBIN, T. R. Contributions to role-taking theory: III. A preface to a psychological analysis of the self. *Psychol. Rev.*, 1952, 59, 11-22.
- TRYON, R. C. *Cluster analysis*. Ann Arbor: Edwards, 1939.

(Received July 13, 1960)

## MANIFEST ANXIETY AND ACHIEVEMENT TEST PERFORMANCE

HAYNE W. REESE

*University of Buffalo*

According to Spence's (1956) theory the effect of motivational or drive level on performance depends on the complexity of the task. In a "simple" task, in which the correct response tendency is stronger than competing response tendencies, subjects with higher drive perform at higher levels than subjects with lower drive; but in a "complex" task, in which the correct response tendency is weaker than competing response tendencies, they perform at lower levels than subjects with low drive. These effects depend on the energizing property of drive. Castaneda (1956) has suggested that the inverse relationship between drive level and performance may also be obtained in simple tasks, such as a reaction time situation, in which the correct response is mediated by implicit verbal responses. Theoretically, a drive produces "drive stimuli" which may be conditioned to responses. If these drive-produced responses interfere with the implicit verbal responses which mediate the instrumental act, the task becomes a complex one and there will be more interference under high drive, since the drive-produced responses have a greater probability of occurrence, than under low drive.

The Children's Manifest Anxiety Scale (CMAS) was designed by Castaneda, McCandless, and Palermo (1956) as a measure of drive in children. Relationships have been demonstrated between CMAS scores and school achievement (McCandless & Castaneda, 1956) and performance on simple and complex learning tasks (Castaneda, Palermo, & McCand-

less, 1956; Palermo, Castaneda, & McCandless, 1956) which indicate that children with high CMAS scores are superior to low scoring subjects on simple tasks and inferior on complex tasks, in line with the interpretation of CMAS scores as indicants of drive level.

It has been suggested that the inverse relationship between CMAS scores and performance on complex tasks may result from the negative correlation between CMAS scores and intelligence which has been found (Hafner & Kaplan, 1959; McCandless & Castaneda, 1956; Phillips, King, & McGuire, 1959). Phillips et al. (1959) reported that when the effects of intelligence were partialled out, there was no relationship between CMAS scores and performance on a battery of psychometric tests differing in "complexity."

The present study was designed to provide additional evidence regarding the relationships among manifest anxiety, intelligence, and performance. An arithmetic test was given under a time limit, and it was assumed that for fourth and sixth grade school children the test would be more nearly a speed test than a power test, i.e., that it would be a simple task in which speed was emphasized (as in the reaction time situation). It was further assumed that the motor responses (writing the answers) would be verbally mediated, and that manifest anxiety would produce responses which would interfere with the mediators (i.e., the symbolic processes). Therefore, the inverse relationship be-

tween manifest anxiety and performance was predicted.

### METHOD

*Subjects.* The subjects were 539 fourth and sixth grade school children from the Williamsville, New York, school district.<sup>1</sup> The data of an additional 85 subjects were eliminated: 54 subjects missed one of the tests, and IQs were not available for 31 subjects.

*Tests and Procedure.* The subjects were given the CMAS and one week later an arithmetic test consisting of 40 addition problems, each with four three-digit numbers. The arithmetic test had a 13-minute time limit. Both tests were administered by the teachers to their regular classes, using standard instructions. Scores on the Primary Mental Abilities Test, which is routinely administered in the third and sixth grades, were obtained from school records.

### RESULTS

Table 1 presents mean scores on the variables, and provides further information about the sample and the difficulty level of the achievement test. (Difficulty level is not necessarily related to "complexity" of the test, since the correct response may be dominant in the response hierarchy but have a low probability of occurrence and therefore a long latency. Under these conditions low scores would result when the time limit is relatively short.)

Correlations among the variables are presented in Table 2.<sup>2</sup> Manifest

<sup>1</sup>The writer is indebted to W. E. Keller, Supervising Principal, Williamsville Central Schools; L. R. Cofran, Principal, Academy Elementary School; R. A. Wilson, Principal, Dodge Elementary School; K. L. Merrick, Principal, Forest Elementary School; and E. Szado, Principal, Maple Elementary School, and their teaching staffs, for their fine cooperation in providing facilities and for their assistance in many other ways. The achievement test was constructed, and the instructions standarized, by P. R. Cutting, Buffalo, New York, whose assistance is gratefully acknowledged.

<sup>2</sup>Castaneda (1956) suggested that there may be two kinds of anxiety, "verbal" and "nonverbal" or "somatic." The former should

TABLE 1  
MEAN IQ, CMAS SCORES, AND  
ARITHMETIC TEST SCORES

Measure	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
<i>N</i>	139	144	118	138
IQ	112.4	108.6	112.3	108.7
Anxiety Score	15.8	14.4	14.9	13.3
Lie Score	5.0	4.3	3.4	3.5
Number Right on Arithmetic Test	11.8	10.9	21.9	20.6
Number Wrong on Arithmetic Test	6.2	6.8	4.3	4.2

anxiety was negatively correlated with the number of correct answers on the achievement test, though the correlation was not significant for fourth grade males. To provide further information about the relationship between manifest anxiety and achievement, the subjects of each sex in each grade were divided into six subgroups on the basis of CMAS scores, and their achievement test scores were compared. The main effect of Anxiety Levels was significant ( $F = 6.73$ ;  $df = 5,464$ ;  $p < .001$ ), but this factor did not interact significantly with grade or sex, indicating that there was a significant relationship between manifest

produce responses which interfere with mediators in the present type of task, but the latter should not. Verbal and somatic subscales were constructed from the CMAS, and it was predicted that the correlations with achievement would be negative for the former and positive for the latter, since subjects with high drive should be superior to subjects with low drive unless the drive produces responses which interfere with the mediators. The average correlations with achievement were  $-.20$  for verbal anxiety and  $-.15$  for somatic anxiety. The failure to confirm the prediction may reflect a lack of validity of the subscales or a flaw in the theoretical analysis.

TABLE 2  
PEARSON PRODUCT-MOMENT CORRELATIONS  
AMONG CMAS SCORES, IQ, AND NUM-  
BER RIGHT ON ACHIEVEMENT TEST

Variables Correlated	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
IQ × R	.46***	.67***	.58***	.48***
Anxiety × R	-.33***	-.15	-.22**	-.19*
Lie × R	-.11	-.22***	-.33***	-.10
Anxiety × IQ	-.23***	-.01	-.18*	.03
Lie × IQ	-.14	-.35***	-.32***	-.17*
Anxiety × Lie	.00	-.24***	-.05	-.23***

Note.—"R" refers to Number Right on the achievement test.

\*  $p < .05$ .  
\*\*  $p < .025$ .  
\*\*\*  $p < .01$ .

TABLE 3  
PARTIAL CORRELATIONS BETWEEN NUMBER  
RIGHT AND MANIFEST ANXIETY, WITH  
IQ PARTIALED OUT AND WITH LIE  
SCORE PARTIALED OUT

Variable Partialed Out	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
IQ	-.26	-.20	-.14	-.23
Lie Score	-.33	-.22	-.25	-.22

anxiety and performance on the achievement test, but the trends were not significantly different in the different sexes or grades. The eta for the combined sexes and grades was about .21, which was about the same as the mean of the Pearson correlation coefficients between anxiety and achievement (cf. Table 2). Therefore, it was concluded that the relationship between anxiety and achievement is negative and linear.

Examination of Table 2 shows that manifest anxiety was more strongly correlated with achievement test

scores than with IQ. Table 3 presents partial correlations between anxiety and achievement, with IQ partialed out, and with lie score partialed out. Partialing out IQ tended to reduce the correlations for females by about .07, and to increase the correlations for males by about .04. Partialing out the lie score increased the correlations an average of about .03.

The multiple correlations for predicting performance from anxiety and IQ were about .04 higher than the correlations for IQ alone; and the combination of anxiety and lie score improved prediction, compared with anxiety alone, by an average of about .07. The use of anxiety, IQ, and lie score improved prediction, compared with IQ alone, by an average of about .04, and compared with anxiety and IQ, by only about .01.

The lie score was significantly negatively correlated with achievement in fourth grade males and sixth grade females. Partialing out IQ reduced the correlations to .03 and -.19, respectively.

## DISCUSSION

The data indicate that the relationship between manifest anxiety and performance on a task requiring mediation is monotonic and negative, but that although statistically significant, it is not a strong relationship. Partialing out IQ had little effect on the correlations, but on the other hand the combination of manifest anxiety with IQ did not appreciably increase the accuracy of predicting performance over that for prediction on the basis of IQ alone. Therefore, it must be concluded that although the relationship between manifest anxiety and performance is not a result of the relationship between anxiety and intelligence, manifest anxiety has relatively little influence on scores on an

achievement test of the type used in the present study.

The lie scale does not seem to serve any useful purpose as far as prediction of performance on the present type of task is concerned, though it does appear to measure a tendency to falsify answers, at least in males.

### SUMMARY

Fourth and sixth grade children were given the Children's Manifest Anxiety Scale and an arithmetic achievement test, consisting of 40 addition problems, which was given with a time limit. It was found that manifest anxiety was inversely and monotonically related to the number of correct responses on the achievement test. Partialing out IQ had little effect on the correlations between manifest anxiety and achievement, but predictions of achievement were not appreciably improved by the combination of manifest anxiety with IQ.

### REFERENCES

CASTANEDA, A. Reaction time and response amplitude as a function of anxiety and

- stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225-228.  
CASTANEDA, A., McCANDLESS, B. R., & PALERMO, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, **27**, 317-326.  
CASTANEDA, A., PALERMO, D. S., & McCANDLESS, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, **27**, 327-332.  
HAFNER, A. J., & KAPLAN, A. M. Children's manifest anxiety and intelligence. *Child Develpm.*, 1959, **30**, 269-271.  
McCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, **27**, 379-382.  
PALERMO, D. S., CASTANEDA, A., & McCANDLESS, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, **27**, 333-337.  
PHILLIPS, B. N., KING, F. J., & MCGUIRE, C. Studies on anxiety: I. Anxiety and performance on psychometric tests varying in complexity. *Child Develpm.*, 1959, **30**, 253-259.  
SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

(Received July 25, 1960)

## CREATIVE AND ACADEMIC PERFORMANCE AMONG TALENTED ADOLESCENTS<sup>1</sup>

JOHN L. HOLLAND<sup>2</sup>

*National Merit Scholarship Corporation, Evanston, Illinois*

The need to understand the nature of academic achievement and creative behavior is pressing; such knowledge is vital for the proper development of educational practice, the administration of scholarship and fellowship programs, and the fostering of talented behavior generally. Unless attempts to encourage the development of creative talent rest on valid assumptions about the nature of academic and creative behavior, such efforts will have little useful influence.

The present study is one of several investigations planned to explore the nature of academic achievement and creative behavior in adolescents and young adults (Holland, 1959a, 1959b, 1960a). The earlier studies were concerned primarily with academic performance. This study tests a large number of hypotheses about variables which are often assumed to be associated with both academic and creative achievement—aptitude scores, self and teacher ratings, originality measures, parental attitudes and values, vocational interests and aspirations, and background information. Most of the variables were used to test a limited number of hypotheses; a few were included for exploratory purposes. The hypotheses about the nature of the academic achiever were derived in part from an extensive lit-

erature which suggests that the student who gets good grades in high school is bright, persistent, conforming, self-controlled, responsible, serious, and rated high by his teachers (d'Heurle, Mellinger, & Haggard, 1959; Gough, 1953; Holland, 1959a, 1959b). Drews and Teahan (1957) indicate that the academic achiever tends to have a somewhat authoritarian mother. Several recent reports (Holland, 1960a; MacKinnon, 1959) imply also that good grades in high school and college may be either unrelated to, or negatively correlated with, potential for creative performance. Taken together, these findings have somewhat consistent implications about the academic achiever and his parents' attitudes and values. The hypotheses derived from the literature and tested here are summarized as follows:

1. Outstanding academic achievement—High School Rank (HSR)—will be positively associated with high scores on the Scholastic Aptitude Test (Math and Verbal factors) (SAT-M and SAT-V), and on the Mastery, Deferred Gratification, and Control scales; with high Self-Ratings of Drive to Achieve and Perseverance; with high Teacher Ratings of Maturity, Popularity, and Social Leadership; with Fathers' Values of Good Student; and with mothers' authoritarian attitudes as measured by the Parental Attitude Research Inventory (PARI).

2. Outstanding academic achievement will be associated with low scores on measures of creativity and originality (Differential Reaction Sched-

<sup>1</sup> This study was partially supported by research grants from the National Science Foundation and the Old Dominion Foundation.

<sup>2</sup> I am indebted to the following staff members for their skillful assistance in all phases of this project: Virginia Chalmers, Laura Kent, Donald Thistlethwaite, and Elizabeth van Laer.

ule, Complexity-Simplicity, Independence of Judgment, and Barron Originality scales) and with low self-evaluations.

The hypotheses about the nature of the creative student were drawn principally from the work of Barron (1953a, 1953b), Gough (1957a), and MacKinnon (1959), which depicts the creative person as independent, complex in outlook, curious, self-assured, intellectual, interested in science and art, and generally effective. This characterization implies also a more permissive parental background which encourages independence and the exploration of self and environment. The hypotheses derived from this literature are summarized as follows.

1. Creative performance (winning public competitions which demand creativity) will be positively associated with high scores on measures of creativity and originality (Differential Reaction Schedule, Complexity-Simplicity, Independence of Judgment, and Barron Originality scales), and on the Initiative and Self-Assurance scales; with high Self-Ratings of Originality, Independence, and Perseverance; with Fathers' Values of Curious and Independent; and with mothers' nonauthoritarian attitudes.

2. Creative performance will have no correlation, or negative correlations, with good high school grades; with high Self-Ratings of Popularity; with high Teacher Ratings; and with Fathers' Values of Dependable, Good Student, Happy and Well-Adjusted, and Popular.

#### METHOD

##### *Student and Parent Sample*

The student and parent samples were obtained from a one-sixth random sample of National Merit Finalists (approximately 9,868 high-scoring students by state from an initial pool of 478,991 high school juniors),

who were polled and tested by mail. An 84% return was obtained, but incomplete information reduced the sample to 59%, 649 boys and 345 girls, and their parents. The average aptitude level of these samples on the SAT were: For boys, SAT-V and SAT-M, 658.4 and 697.9 with SDs of 56.8 and 62.2, respectively. For girls, the SAT-V and SAT-M means are 659.2 and 635.4 with SDs of 60.6 and 70.3.

##### *Academic and Creative Performance*

The criterion of academic performance was high school grades (HSR) during the first 3 years of high school. The criteria of creative performance were derived from a checklist of accomplishments assumed to require creative or original behavior. Creative performance is defined as a performance which is accorded public recognition through awards, prizes, or publication, and which may therefore be assumed to have exceptional cultural value. Because of the difficulty in arriving at a generally acceptable definition of "creativity," these criteria should perhaps be regarded as either "notable scientific or artistic performance," although we will refer to the criterion as "creative" performance hereafter to enhance readability. With this definition as a guide, a list of 20 achievements at the high school level was derived by reviewing the secondary school achievements of Finalists from previous years. Items were divided by content into two scales: Creative Science (5 items) and Creative Arts (11 items). (Four of the original 20 items were omitted because they appeared to be inadequate signs of creative behavior.) The items on the two scales are shown below. Students were asked to check those which applied to them.

##### Creative Science Scale

- 1. Gave an original paper at a scientific meeting sponsored by a professional society.
- 2. Won a prize or award in a scientific talent search.
- 3. Constructed scientific apparatus on own initiative.
- 4. Invented a patentable device.
- 5. Had scientific paper published in a science journal.

Item 3 does not meet the criteria of "public recognition" and "unusualness," but in order to lengthen the scale, it was in-

cluded as a low-level sign of creative performance. Similarly, Item 4 depends on the student's judgment, and may or may not be an accomplishment which has received public recognition.

### Creative Arts Scale

1. Won one or more speech contests.
2. Had poems, stories, or articles published in a *public* newspaper or magazine (not school paper) or in a state or national high school anthology.
3. Won a prize or award in an art competition (sculpture, ceramics, painting, etc.).
4. Received the highest rating in a state music contest.
5. Received one of the highest ratings in a national music contest.
6. Composed music which has been given at least one public performance.
7. Arranged music for public performance.
8. Had minor roles in plays (not high school or church-sponsored).
9. Had leads in high school or church-sponsored plays.
10. Won literary award or prize for creative writing.
11. Had cartoon published in public newspaper or magazine, etc. (not high school paper).

For boys, the estimated reliabilities (Kuder-Richardson Formula 20) of the Arts and Science scales, respectively, are .36 and .55. For girls, the reliabilities are .38 and .37. These low reliabilities may be attributed to the brevity of the scales; they limit the level of correlation between the predictors and these criteria. The Arts and Science scales correlate with one another .15 and .09 for boys and girls, respectively.

### Predictors

Students and parents were sent the following inventories and scales. Students filled out:

1. The Gough (1957b) Differential Re-

action Schedule (DRS), an inventory which was designed to predict originality and which contains the following scales: Intellectual Competence, Inquiringness, Cognitive Flexibility, Esthetic Sensitivity, Sense of Destiny, Total Score (sum of all scores), and Potential Success (PIV), an index of general drive and ambition derived from a variety of Institute of Personality Assessment and Research (IPAR) studies of achievement and personal effectiveness.

2. The Complexity-Simplicity, Independence of Judgment, and Originality scales from Barron's (1952) Inventory of Personal Philosophy. Only 54 of the 150 items in the Originality scale were used. The Complexity-Simplicity scale consists of items from the original 900-item IPAR inventory, selected for their significant correlations with scores on the Barron-Welsh Art Scale in five different samples, four groups of college students and one of military officers. The successive sample method was used for item selection; on cross-validation the final scale correlated .7 with the Figure Preference scale. The Originality scale was developed out of an MMPI-CPI pool by correlating each item with a composite score on four measures of originality (Guilford's Unusual Uses, Consequences, and Plot Titles; and the IPAR Word Rearrangement test) in a sample of 343 military officers. It was cross-validated by correlating it with a composite of the three Guilford tests in a sample of 150 undergraduates at the University of California.<sup>3</sup>

3. The Mastery scale, an 18-item scale derived from three items on Strodtbeck's Value Scale (1958).

4. The Deferred Gratification scale, formerly called the Play scale, from the National Merit Student Survey (Holland, 1960a).

5. The Vocational Preference Inventory (VPI), a short revision of the Holland Vocational Preference Inventory (Holland, 1958).

6. The Ghiselli Self-Description Inventory. Since the correlations between weighted and unweighted scale scores for the Initiative, Self-Assurance, and Occupational Level scales ranged from .89 to .96, the items of these scales were scored with unit weights rather than being weighted differentially.

7. Self-ratings on a four-point scale and a Self-Evaluation score based on the number

---

<sup>3</sup> Frank Barron, personal communication, June 1960.

of times a student rates himself above average on 20 traits.

8. The Creative Activities scale (CAS), a list of hobbies and activities assumed to demand original behavior. Teacher Ratings of Citizenship, Popularity, and Social Leadership were also available for each student, as well as SAT scores, level of degree sought, Birth Order, and a Breadth of Interest score based on the number of activities and interests checked from a list of 47 items.

Mothers filled out the PARI (Schaefer & Bell 1958), and 16 of the 23 five-item scales were used. Fathers gave information on family background and ranked nine goals and traits in the order in which they considered them valuable for their children to possess.

## RESULTS

*Correlational analysis.* The 75 predictors and criteria were intercorrelated (product-moment) and the complete  $75 \times 75$  matrices for boys and girls, along with the means and standard deviations, have been deposited with the ADI.<sup>4</sup> Tables 1 and 2 show only the correlations between the three criteria of academic and creative performance and the 72 predictive variables. The influence of intelligence has been partialled out, using the SAT-M score as an estimate of intelligence for all three criteria, since the math score has higher correlations with the three criteria than does the verbal score. Generally, SAT-M has low correlations with the predictors. The largest

<sup>4</sup>The following tables have been deposited with the American Documentation Institute: Tables A and B, Correlational matrices for boys and girls; Tables C and D, Means and standard deviations for boys and girls on 75 variables; and Tables E and F, Daydreams about future occupations for students who are high and low in academic and creative performance. Order Document No. 6611 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

absolute difference between corrected and uncorrected  $r$ 's was only .05.

For boys, a review of the correlations at the 1% level reveals that the three kinds of performance (scientific, artistic, and academic) are associated with somewhat different variables, generally in expected directions.

The boy with a high score on the Creative Science Scale also has many artistic achievements (3) and Creative Activities (8), plans to get an advanced degree (71), is a first-born or an only child (74), and has high scores on the Independence of Judgment (17), Mastery (19), Deferred Gratification (20), Initiative (23), Physical Activity (28), and Intellectuality (29) scales and low scores on the Responsibility (30) and Status (36) scales. He rates himself high on Originality (37), Independence (40), and Perseverance (42). Fathers of boys high on scientific performance regard curiosity (48) as a valuable trait for their sons to have; mothers of these boys tend to be agreeable (supposedly comfortable in their role as mothers) rather than irritable (59).

The boy with a high score on the Creative Arts Scale engages in many Creative Activities (8), plans to get an advanced degree (71), and has high scores on the Esthetic Sensitivity (12), Sense of Destiny (13), DRS (Total Score) (14), Breadth of Interest (21), Self-Evaluation (22), Self-Assurance (24), Occupational Level (25), Response Bias (Acquiescence) (26), Responsibility (30), Verbal Activity (32), Emotionality (33), Control (34), and Status (36) scales, and low scores on SAT-M (5). He rates himself high on Originality (37), Drive to Achieve (39), Independence (40), Self-Confidence (41), and Perseverance (42). Mothers of these boys are slightly more accepting of the home-making role than are mothers of stu-

J TABLE 1

CORRELATIONS BETWEEN THREE CRITERIA OF ACADEMIC AND CREATIVE PERFORMANCE  
AND 72 APTITUDE, PERSONALITY, AND BACKGROUND  
VARIABLES FOR BOYS  
(N = 649)

Variable	Creative Performance		High School Grades	Variable	Creative Performance		High School Grades
	Scientific	Artistic			Scientific	Artistic	
1. HSR	.00	.08	—	39. Drive to Achieve	.08*	.22**	.28**
2. Scientific Performance	—	.15	.00	40. Independence	.15**	.20**	.06
3. Artistic Performance	.15**	—	.08	41. Self-Confidence	.10*	.23**	.13**
4. SAT-V	.02	-.09*	-.04	42. Perseverance	.11**	.15**	.27**
5. SAT-M	.04	-.19**	.09*	Teacher Ratings			
6. Humanities Comprehension	-.01	.03	.02	43. Citizenship	.00	.09*	.26**
7. Scientific Comprehension	.06	.02	.04	44. Popularity	.01	.05	.17**
8. Creative Activities (NMSS)	.36**	.37**	-.03	45. Social Leadership	.00	.10*	.19**
Gough DRS				Father's Values and Goals			
9. Intellectual Competence	-.03	.02	-.01	46. Defend Self	-.01	.00	-.02
10. Inquiringness	.00	.02	.02	47. Ambitious	.06	.00	.05
11. Cognitive Flexibility	.02	.01	-.09*	48. Curious	.11**	-.08*	-.09*
12. Esthetic Sensitivity	.01	.12**	.01	49. Dependable	-.10*	-.02	.02
13. Sense of Destiny	.03	.23**	-.04	50. Good Student	-.02	.06	-.06
14. Total Score (Above 5)	.03	.15**	-.03	51. Happy, Well-Adjusted	-.02	.01	-.12**
15. Potential Success (PIV)	.01	.04	.01	52. Independent, Self-Reliant	-.02	.04	.09*
Originality-Personality				53. Popular	-.01	.01	
16. Complexity-Simplicity	.07	.07	-.14**	54. Self-Controlled	-.01	.00	-.06
17. Independence of Judgment	.10*	.03	-.11**	PARI			
18. Barroo Originality	.07	.08*	-.12**	55. Fostering Dependency	-.01	-.06	.05
19. Mastery	.11**	.10*	.13**	56. Seclusion of Mother	.01	-.02	.10*
20. Deferred Gratification	.11**	.06	.11**	57. Martyrdom	.03	-.03	.09*
21. Breadth of Interest	.09*	.23**	.00	58. Strictness	-.07	.03	-.02
22. Self-Evaluation	.01	.23**	.24**	59. Irritability	-.11**	-.05	.06
Ghiselli				60. Excluding Outside Influences	.03	-.06	
23. Initiative	.11**	.07	-.02	61. Suppression of Aggression	.07	-.04	.08*
24. Self-Assurance	.09*	.12**	.11**	62. Rejection of Homemaking Role	-.02	-.11**	.02
25. Occupational Level	.01	.11**	-.07	63. Equalitarianism	-.05	.03	.02
VPI				64. Approval of Activity	.00	-.07	.06
26. Response Bias	-.02	.13**	.05	65. Avoidance of Communication	.05	.03	.12**
27. Infrequency	-.07	-.01	.03	66. Suppression of Sex	.03	-.04	.05
28. Physical Activity	.18**	-.04	.02	67. Ascendancy of Mother	-.05	-.09*	.05
29. Intellectuality	.23**	.03	.07	68. Intrusiveness	.10*	.03	.05
30. Responsibility	-.17**	.20**	.05	69. Comradeship and Sharing	-.07	-.02	.06
31. Conformity	-.07	.01	.06	70. Acceleration of Development	.07	-.03	.10*
32. Verbal Activity	-.10*	.11**	.00	Miscellaneous			
33. Emotionality	-.05	.31**	.02	71. Degree (level) Sought	.15**	.13**	.04
34. Control	-.02	.12**	.15**	72. Father's Educational Level	.05	.05	-.08*
35. Masculinity-Femininity	.10*	-.05	.03	73. Mother's Educational Level	-.02	.03	-.10*
36. Status	-.18**	.18**	.12**	74. Birth Order (first born)	.09*	-.04	.01
Self-Ratings				75. School Size	-.02	-.04	.03
37. Originality	.15**	.27**	-.01				
38. Popularity	-.05	.10*	.12**				

\* Significant at the .05 level.

\*\* Significant at the .01 level.

TABLE 2

CORRELATIONS BETWEEN THREE CRITERIA OF ACADEMIC AND CREATIVE PERFORMANCE  
AND 72 APTITUDE, PERSONALITY, AND BACKGROUND VARIABLES FOR GIRLS  
(N = 345)

Variable	Creative Performance		High School Grades	Variable	Creative Performance		High School Grades
	Scientific	Artistic			Scientific	Artistic	
1. High School Grades	.00	-.09	—	39. Drive to Achieve	15**	.13*	.18**
2. Scientific Performance	—	.09	-.03	40. Independence	.07	.12*	-.04
3. Artistic Performance	.09	—	-.09	41. Self-Confidence	.05	.11*	.12*
4. SAT-V	.07	.01	.08	42. Perseverance	.13*	.13*	.22**
5. SAT-M	.13*	-.12*	.24**				
6. Humanities Comprehension	.03	.05	.07	Teacher Ratings			
7. Scientific Comprehension	.11*	-.01	.05	43. Citizenship	.05	-.04	.30**
8. Creative Activities (NMSS)	.23**	.36**	-.19**	44. Popularity	-.02	-.04	.23**
Gough DRS				45. Social Leadership	.01	.04	
9. Intellectual Competence	-.14**	-.02	-.01	Father's Values and Goals			
10. Inquiringness	.00	.02	-.03	46. Defend Self	.08	.00	.00
11. Cognitive Flexibility	.00	.09	-.12*	47. Ambitious	.13*	.05	.04
12. Esthetic Sensitivity	.06	.06	-.01	48. Curious	.06	.03	-.16**
13. Sense of Destiny	.03	.23**	-.04	49. Dependable	-.15**	-.02	.16**
14. Total Score (Above 5)	-.01	.14**	-.08	50. Good Student	.11*	.08	-.01
15. Potential Success (PIV)	-.04	.00	.00	51. Happy, Well-Adjusted	-.13*	.00	-.02
Originality-Personality				52. Independent, Self-Reliant	.00	-.09	-.01
16. Complexity-Simplicity	.02	.06	-.14**	53. Popular	-.11*	-.04	.00
17. Independence of Judgment	.06	.02	-.11*	54. Self-Controlled	-.05	-.01	.02
18. Barron Originality	.07	.10	-.05				
19. Mastery	.19**	.02	.08	PARI			
20. Deferred Gratification	.11*	.03	.14**	55. Fostering Dependency	-.02	-.10	.04
21. Breadth of Interest	.18**	.24**	-.09	56. Seclusion of Mother	-.03	.02	.15**
22. Self-Evaluation	.08	.19**	.13*	57. Martyrdom	-.06	-.05	.08
Ghiselli				58. Strictness	.03	-.01	.07
23. Initiative	-.03	.07	.03	59. Irritability	-.08	-.07	.03
24. Self-Assurance	.03	.00	.15**	60. Excluding Outside	-.01	-.10	.11*
25. Occupational Level	.01	.10	.01	Influences			
VPI				61. Suppression of Aggression	-.06	-.06	.16**
26. Response Bias	.00	.05	.09	62. Rejection of Homemaking	-.06	-.05	.06
27. Infrequency	-.05	-.06	.06				
28. Physical Activity	.08	-.04	-.03	63. Equalitarianism	-.04	-.06	.01
29. Intellectuality	.11*	-.07	-.06	64. Approval of Activity	-.01	-.06	.13*
30. Responsibility	.02	.06	.06	65. Avoidance of Communica-	.00	-.07	.13*
31. Conformity	-.03	-.07	.02	tion			
32. Verbal Activity	-.06	.06	.11*	66. Suppression of Sex	-.05	-.06	.10
33. Emotionality	-.02	.10	.07	67. Ascendancy of Mother	-.01	-.13*	.05
34. Control	.04	-.01	.14**	68. Intrusiveness	-.09	-.03	.05
35. Masculinity-Femininity	.06	.06	-.02	69. Comradeship and Sharing	-.08	-.13*	.00
36. Status	-.02	.02	.16**	70. Acceleration of Develop-	.04	-.01	.15**
Self-Ratings				ment			
37. Originality	.13*	.32**	-.03	Miscellaneous			
38. Popularity	-.01	.10	.04	71. Degree (level) Sought	.19**	.09	-.03
				72. Father's Educational Level	.06	.10	-.08
				73. Mother's Educational	.01	.13*	.00
				Level			
				74. Birth Order (first born)	-.08	-.02	-.01
				75. School Size	-.03	-.05	.02

\* Significant at the .05 level.

\*\* Significant at the .01 level.

dents with low artistic performance (62).

The variables related to academic performance are somewhat different from those related to creative performance. Academic performance (HSR) is negatively correlated with high scores on the Complexity-Simplicity (16), Independence of Judgment (17), and Barron Originality (18) scales and positively correlated with high scores on the Mastery (19), Deferred Gratification (20), Self-Evaluation (22), Self-Assurance (24), Control (34), and Status (36) scales. Boys with high grades rate themselves high on Popularity (38), Drive to Achieve (39), Self-Confidence (41), and Perseverance (42). Their teachers rate them high on Citizenship (43), Popularity (44), and Social Leadership (45). Fathers of these students want their sons to be Good Students (50) and are less concerned with their being Independent and Self-Reliant (52). Mothers tend to be authoritarian (at the 1% level), Avoidance of Communication (65); at the 5% level, Seclusion of Mother (56), Martyrdom (57), Suppression of Aggression (61), and Acceleration of Development (70).

The findings for girls, shown in Table 2, are similar to those for boys. The girl with a high score on the Creative Science Scale tends to engage in many Creative Activities (8), plans to get an advanced degree (71), has low scores on the Intellectual Competence scale (9) and high scores on the Mastery (19), Breadth of Interest (21) scales. She rates herself high on Drive to Achieve (39). Her father ranks being Dependable (49) which he values for his children.

The girl with a high score on the Creative Arts Scale has high scores on the Creative Activities (8), Sense

of Destiny (13), DRS (Total Score) (14), and Breadth of Interest (21) scales; has a positive Self-Evaluation (22), and rates herself high on Originality (37).

Academic performance is negatively correlated with high scores on the Creative Activities (8) and Complexity-Simplicity (16) scales and positively correlated with high scores on SAT-M (5) and the Deferred Gratification (20), Control (34), Status (36), and Self-Assurance (24) scales; with high Self-Ratings of Drive to Achieve (39) and Perseverance (42), and high Teacher Ratings of Citizenship (43), Popularity (44), and Social Leadership (45). Fathers of girls who get high grades value being Dependable (49) and do not value being Curious (48). Mothers of these girls are characterized by the PARI as "authoritarian"—Seclusion of Mother (56), Suppression of Aggression (61), Acceleration of Development (70).

The differences between creative and academic performance for the 72 variables replicate the trends found for boys. Academic performance appears to be the function of a personal syndrome characterized by perseverance, self-control, good behavior (good citizenship), and rigidity; this personality pattern is related to parental attitudes which seem conducive to such a pattern. Creative performance, on the other hand, seems to be the outcome of a conscious conception of being original, active participation in creative hobbies, and reinforcement by parents who possess values and attitudes which appear to be conducive to such performance.

*Free response analysis.* As a part of the assessment, students were asked about their vocational choice, the factors which influenced this choice, and their images and daydreams about their future occupations. Their free

responses to the question, "What do you daydream about when you think of your future occupation?" were coded, using 10 coding categories developed from a sample of 200 students and tested for reliability on a sample of 160 students. Two raters, coding independently, obtained 75% agreement. Tables E and F<sup>4</sup> compare the responses of students high on creative performance with those of students low on creative performance, and of students high on academic performance with students low on academic performance.

According to Table E, boys with outstanding scientific performance daydream of "high achievement and creative accomplishment"; for this scoring category, the percentage difference between high and low science scores is significant at the .01 level. The same trend is observable in the Arts scale, though the percentage difference is not significant at the .05 level. The responses of students above the median on HSR do not differ significantly from those of students below the median.

Girls with outstanding scientific achievement daydream more about work activities than about helping others ( $p < .001$ ), whereas girls with low achievement are less concerned with work and more interested in helping others ( $p < .05$ ). Again, the comparisons of girls high and low on academic achievement do not reveal significant differences.

These results, which support some of the correlational findings, imply that academic achievement involves somewhat different motives than creative performance; good grades in high school appear to be a function of socialization (citizenship and popularity) and perseverance, whereas creative performance is a function of conscious concern with high accom-

plishment, independence, and originality.

*Getzel's analysis.* In a recent study of elementary school children, Torrance (1959), following Getzel's classification of students as "high IQ" or "high creative," reports that the most intelligent students (top 20%) are not necessarily the most creative (top 20%); only about 30% of the students were in both groups.

The present study supports Torrance's findings. Eight tables were obtained for the present study by relating SAT-V and SAT-M to the scientific and artistic checklist scores for each sex. Within each table, distributions of creative performance scores were formed for approximately the top 20% of students on the SAT and the top 20% of students on creative performance. The percentages of overlap between these two contrasting student groups ranged from 1-6.5% of the total samples for the eight tables and deviate only slightly from an expected overlap of 4%. The larger deviations were in the direction expected from the correlations in Tables 1 and 2. These results are consistent with the negligible correlations between aptitude and outstanding creative achievement shown in Tables 1 and 2. We may conclude, then, that in a group of exceptionally bright students, intelligence has little or no relationship to creative performance in arts and science or to academic achievement.

These results may in part be a function of the regression of creativity on aptitude; students are selected initially for their high aptitude, and then the relationships between aptitude and creativity are studied within this narrow range of aptitude. These low relationships are, only in part, a function of the narrow range on aptitude, however. In Tables A and B of the ADI material SAT-V and SAT-M

correlate as much as .64 and .57 with the Humanities and Scientific Comprehension scores of the National Merit Scholastic Scholarship Test (an achievement test given about one year before the SAT was administered). Even these correlations are probably underestimates, due to the artificial restriction of range imposed on the NMSQT by its use as a selection device for this sample. The latter findings suggest that the SAT does discriminate efficiently and reliably within this narrow range of talent so that higher relationships between aptitude and creative as well as academic performance were possible.

### DISCUSSION

The results of the present study are probably conservative. The distribution of scores for all criteria are at least moderately skewed and all scales have relatively low reliability; moreover, the range of scores on the remaining predictor variables, including background variables, is restricted. Therefore, the correlational results are probably underestimated. Since the students in the sample are all exceptionally talented, the correlations between originality measures and scholastic aptitude are reduced. Whether or not the observed relationships can be generalized to more representative student populations is questionable.

The correspondences between the present study and related studies are numerous. The originality scales developed by Barron (Independence of Judgment, Complexity-Simplicity, and Originality) and by Gough (DRS) have been frequently correlated with criteria of creative performance and our results are generally consistent with earlier reports (Barron, 1953a, 1953b; Gough, 1957b). Although the relationships obtained in this study

are generally low, it is surprising that the Barron and Gough scales correlate at all, since they were constructed for adult samples and used different criteria of creative behavior.

The CAS has promising validity since it is the best single correlate of creative performance ( $r$ 's range from .23-.37 for the CAS against the Arts and Science criteria for both sexes), and since its highest correlations are with a set of variables which lend support to this interpretation. For boys, the variables having the largest correlations with the CAS are, in descending order (.37-.15): Breadth of Interest, Self-Ratings of Originality and Independence, Emotionality, Self-Ratings of self-control, Total DRS Score, Complexity-Simplicity, level of degree sought, and Barron Originality. For girls, the largest correlates of the CAS are (.43--.16): Breadth of Interest, Self-Ratings of Originality and Independence, Barron Originality, Complexity-Simplicity, Sense of Destiny, Self-Evaluation, and mother's education. The Fostering Dependency and Ascendancy of Mother scales of the PARI correlate negatively with the CAS (-.16 and -.17, respectively).

The assumed validity of the CAS may be overestimated, since 4 or 5 of the 32 scale items involve activities which are preparatory to or associated with the accomplishments listed in the Creative Science and Arts Scales. The correlates of the CAS listed above imply some validity independent of the CAS's correlations with the performance scales, however.

The CAS was suggested by the Chorness and Nottelmann study (1957) in which it was found that a set of extracurricular hobbies had significant correlations (.27-.45) with four or five Guilford factor composites when intelligence is partialled out. The item overlap, if any, between the

Chorness creative hobbies and the CAS is not known.

The unexpected concurrent validity of the Self-Ratings is of special interest; of all the variables used, the Self-Ratings show most clearly that academic and creative performers conceive of themselves as different from other people. It is of interest too that high Self-Ratings of Perseverance and Self-Confidence—two of the three most discriminating traits for Terman's high and low achievers—are correlated here with both creative and academic performance (Terman & Oden, 1947).

The findings for the Intellectuality and Emotionality scales of the VPI appear to replicate several studies (Barron, unpublished; MacKinnon, 1959) in which vocational interests in science and the arts were found to be associated with rated creative performance and high scores on tests of creative behavior. Since the Intellectuality scale has its highest significant correlations with the Self-Sufficiency, Schizothymia, and Masculinity scales of the 16 PF, the present results are consistent with the characterization of the creative person as independent, asocial, and masculine. Similarly, the Emotionality scale of the VPI is correlated most highly with femininity, instability, and introversion (16 PF) (Holland, 1960b).

Generally, the results seem to support many other hypotheses about the nature of the creative person, but an attempt to relate all of the specific findings to the voluminous theorizing about creativity would be a gargantuan task. Briefly, those hypotheses which regard the creative person as independent, intellectual, expressive, asocial, consciously original, and open to experience gain some support (Taylor, 1955, 1957, 1959). The creative student not only has an identi-

fiable personal disposition, but also he has relationships with teachers and parents which are in accordance with our expectations. His parents appear to be more permissive and more nurturant of his ideas and impulses so that communication with the self and the world is stimulated. It is not surprising, then, that such students appear to come in conflict with teachers who demand conforming, controlled, nonexploratory behavior.

Perhaps the most unequivocal finding in the present study is that, for samples of students of superior scholastic aptitude, creative performance is generally unrelated to scholastic achievement and scholastic aptitude. Since the traditional predictors of scholastic aptitude are of little or no value for predicting creativity, it seems clear that scholarship programs, colleges, and other agencies, if they are concerned primarily with rewarding students or selecting employees who have potential for creative performance, need to make a more active effort to devise predictors of creative potential. In fact, attempts to build better scholastic aptitude tests may even be detrimental, since they may lead to a greater dependence on instruments which are of limited value and thus delay unnecessarily the development of efficient predictors of creative performance.

The finding that teachers generally rate students with good grades higher than those with lower grades, together with the results of an earlier study (Holland, 1959b) which indicates that teachers give lower ratings to students with more potential for creativity (as measured by the 16 PF) than to students with less potential, suggests that school personnel value the good grade-getter more than the creative student. Recently, Jex (1958) found that school principals rate teachers with high

scores on an ingenuity test lower than teachers with low test scores. Taken together, these results reinforce a common belief that some of our traditional education may be stultifying rather than nurturant and fructifying.

### SUMMARY

The relationships between three criteria of academic and creative performance and 72 personal, demographic, and parental variables were studied in a sample of talented adolescents. The results suggest that creative performance at the high school level occurs more frequently among students who are independent, intellectual, expressive, asocial, consciously original, and who have high aspirations for future achievement. Students who are persevering, sociable, responsible, and whose parents hold somewhat authoritarian attitudes and values, are more frequently academic achievers. The negligible relationships found between academic aptitude and creative performance at a high aptitude level suggest that we need to use nonintellectual criteria in the selection of students for scholarships and fellowships.

### REFERENCES

- BARRON, F. *Inventory of personal philosophy*. Berkeley, Calif.: Univer. California Press, 1952. (Institute of Personality Assessment and Research)
- BARRON, F. Complexity-simplicity as a personality dimension. *J. abnorm. soc. Psychol.*, 1953, **48**, 163-172. (a)
- BARRON, F. Some correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287-297. (b)
- CHORNES, M. H., & NOTTELmann, D. N. The prediction of creativity among Air Force civilian employees. *USAF Personnel Train. Res. Cent. res. Rep.*, 1957, No. AFPTRC-TN57-36.
- D'HEURLE, ADMA, MELLINGER, JEANNE C., & HAGGARD, E. A. Personality, intellectual, and achievement patterns in gifted children. *Psychol. Monogr.*, 1959, **73** (13, Whole No. 483).
- DREWS, ELIZABETH M., & TEAHAN, J. E. Parental attitudes and academic achievement. *J. clin. Psychol.*, 1957, **13**, 328-332.
- GOUGH, H. G. What determines the academic achievement of high school students? *J. educ. Res.*, 1953, **46**, 321-331.
- GOUGH, H. G. Imagination: Undeveloped resource. Paper read at Proceedings of the first conference on research and development in personnel management, Los Angeles, California, 1957. (a)
- GOUGH, H. G. Researchers' summary for the Differential Reaction Schedule. Univer. Calif., 1957. (Ditto) (b)
- HOLLAND, J. L. A personality inventory employing occupational titles. *J. appl. Psychol.*, 1958, **42**, 336-342.
- HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *J. educ. Psychol.*, 1959, **50**, 135-142. (a)
- HOLLAND, J. L. Some limitations of teacher ratings as predictors of creativity. *J. educ. Psychol.*, 1959, **50**, 219-223. (b)
- HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245-254. (a)
- HOLLAND, J. L. The relation of the Vocational Preference Inventory to the Sixteen Personality Factor Questionnaire. *J. appl. Psychol.*, 1960, **44**, 291-296. (b)
- JEX, F. B., & MERRILL, R. M. An evaluation of the first academic-year institute, University of Utah, 1957-58. *U. Utah res. Monogr. Educ.*, 1958, **1**, No. 2.
- MACKINNON, D. W. What do we mean by talent and how do we test for it? In, *The search for talent: College admissions 7*. New York: College Entrance Examination Board, 1959.
- SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Develpm.*, 1958, **29**, 339-361.
- STRODTBECK, F. L. Family interaction, values, and achievement. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and society*. New York: Van Nostrand, 1958. Ch. 4.
- TAYLOR, C. W. (Ed.) *The first, second, and third University of Utah research con-*

- ferences on the identification of creative scientific talent. Salt Lake City: Univer. Utah Press, 1955, 1957, 1959.
- TERMAN, L. M., & ODEN, MELITA H. *Genetic studies of genius: IV. The gifted child grows up*. Stanford: Stanford Univer. Press, 1947.
- TORRANCE, E. P. Explorations in creative

thinking in the early school years: A progress report. In, *The third University of Utah research conference on the identification of creative scientific talent*. Salt Lake City: Univer. Utah Press, 1959. Pp. 58-71.

(Received September 26, 1960)

# BEHAVIORS DURING PROBLEM SOLVING AMONG CHILDREN OF LOW, AVERAGE, AND HIGH INTELLIGENCE<sup>1</sup>

HERBERT J. KLAUSMEIER  
*University of Wisconsin*

AND

LEO J. LOUGHLIN  
*Northern Illinois University*

The range of differences among school children in many intellectual and psychomotor abilities and other characteristics is defined clearly. Less clearly defined are the factors which contribute to the differences. Previously, no statistically significant difference was found among groups of low, average, and high IQ children in retention of arithmetic problems, graded to each child's achievement level (Klausmeier & Feldhusen, 1959); nor in emotional adjustment, integration of self-concept, and achievement in relation to expectancy (Klausmeier, Check, & Feldhusen, 1960). Large differences with little overlapping between IQ levels were found in achievements in reading, language, and arithmetic. The purpose of this investigation was to ascertain possible differences among the three groups in efficiency of method, persistence, and mode of attack while in the actual process of solving arithmetic problems.

## PROCEDURE

### Subjects

The subjects were 20 boys and 20 girls with WISC IQs of 55-80, 20 boys and 20 girls with WISC IQs of 90-110, and 20 boys and 20 girls with WISC IQs of 120-146. The 80 children of high and average IQ were enrolled in 15 regular fifth grade classes of two large public school systems; and the 40 low IQ children were enrolled in 14 special classes for educable mentally retarded children in the same schools. Excluded from

the study were low IQ children exhibiting a second handicap to a marked degree or definite organic symptoms of retardation such as mongolism. All had birthdates between September 15, 1947, and December 15, 1948; and the mean age at the time of gathering the data was 131 months for each IQ group.

### Methods

The first task was to get problems properly graded in difficulty. After extensive experimenting with pupils in regular classrooms and with low IQ pupils not in the study, a group of 29 problems at three main levels of difficulty was arranged. The problems for the low IQ children were subdivided further with the lowest level intended for those who did not know the number of cents in a dime or who could not perform simple addition. A reserve level of problems of high difficulty was established for the high IQ child who might solve an assigned problem in less than 3 minutes. The least difficult problem designed for a low IQ child was to make 7¢ with three coins and the most difficult designed for a high IQ child was to make \$14.53 with 16 bills and coins.

Besides this initial formulation of problems, two other measures were used to make certain that the problem was of appropriate difficulty. A survey test was administered to each child to ascertain his knowledge of our money system, related vocabulary, and mathematical skills. Also, if a child completed his assigned problem in less than 3 minutes, the next problem in order of difficulty was assigned.

Each child solved only one problem. This problem and the survey test mentioned previously were administered by a carefully trained examiner; and another trained person observed the entire procedure. From the time the problem and instructions were given the child until the problem was solved, time in seconds and the incidence of the following behaviors of the child were recorded: (a) manipulations of the child were recorded: (a) manipulations of coins by the low IQ children or manipulation of figures

<sup>1</sup> The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

with pencil and paper by the average and high IQ children; (b) engaging in nontask activities or withdrawal which resulted in the examiner's encouraging the child to continue his problem solving; (c) offering an incorrect solution; (d) noting and correcting mistakes independently; (e) verifying the solution by checking the amount of money independently; (f) verifying the solution by checking the number of coins independently; and (g) using a random, logical, or a combination approach to the solution. In random approach the subject merely manipulated or explored the materials without an apparent plan of attack. In logical approach the subject made deliberate manipulations to test successive trials or hypotheses. Some children changed from an apparently random to a logical approach, and vice versa.

Efficiency of the child's method was defined as the ratio of moves (coins or paper-pencil computations made) to the least number needed to achieve the solution. For example, the least efficient subject made 39 moves when only 2 were required, with a resulting efficiency ratio of .051. Any instance of a child's being encouraged by the examiner to cease from a nontask activity or to continue at the task was recorded as the measure of nonpersistence. The mode of attack was observed in the verbal responses of the subject and in his manipulations of coins or figures, as described in the preceding paragraph, *c* through *f*.

The data secured on each subject were tabulated and combined according to the IQ groups. Analysis of variance was used to

compare efficiency of method and chi square to compare incidences of the other behaviors among the three groups.

## RESULTS

The extent to which the problems were of equal difficulty for each IQ group was ascertained by comparing the time in seconds required to reach a solution to the problem. The low IQ group required a mean time of 624.17 seconds, the average group 627.65, and the high group 578.47 seconds. On the basis of a small and insignificant *F* value obtained among the means, it was concluded that the assigned problems were about equally difficult for the three IQ groups.

A summary of the data and the results of the tests of statistical significance of differences among the three IQ groups and between groups are presented in Table 1. Column 1 lists the observed behaviors. Columns 2-4 contain the scores or number of subjects in each IQ group who demonstrated the behavior. Column 5 shows the results of the tests of statistical significance of differences among the three IQ groups. Columns 6-8 give the results of the tests of sig-

TABLE 1  
SUMMARY OF OBSERVED SCORES AND RESULTS OF TESTS OF STATISTICAL SIGNIFICANCE  
OF DIFFERENCES

Behavior	Score or Incidence			Significance of Differences			
	High	Average	Low	Among Groups	Between Groups		
					Hi-Av	Hi-Low	Av-Low
Efficiency of Method (%)	.353	.309	.237	.05 > p > .01	ns*	.05 > p > .01	ns
Nonpersistence (Incidence)	3	10	17	.01 > p > .001	.02 > p > .01	.01 > p > .001	.02 > p > .01
Mode of Attack (Incidence)	10	16	33	.001 > p	.001 > p	.001 > p	.01 > p > .001
Offered Incorrect Solution	14	12	5	ns	—	—	—
Noted and Corrected Mistakes	14	12	5	ns	—	—	—
Verified Solution—Amt. Money	38	31	11	.001 > p	.001 > p	.001 > p	.001 > p
Verified Solution—No. Coins	36	26	9	.001 > p	.001 > p	.001 > p	.001 > p
Random Approach	1	8	21	.001 > p	.01 > p > .001	.001 > p	.001 > p
Logical Approach	14	5	3	.01 > p > .001	.01 > p > .001	.01 > p > .001	.05 > p > .02

\* Not significant at .05 level.

nificance of differences between groups—high and average, high and low, and average and low.

Table 1 shows that the low IQ group was inferior to the high IQ and to the average IQ group in seven and six of the observed behaviors, respectively; the average IQ group was inferior to the high IQ in six behaviors. Although the mean number of seconds to solve was nearly identical for the three groups, the incidences of behaviors were greatly different. To clarify the differences, short sketches of the procedures used by a high and a low IQ boy are now given.

Richard had a WISC IQ of 138 and used 585 seconds to solve the problem: "Make \$9.77, using 12 bills and coins." He studied the situation and during the first 3 minutes made a first attempt at solution, getting the amount correct but using 13 instead of 12 bills and coins—9 one dollar bills, 1 half-dollar, 1 quarter, and 2 pennies. He checked the amount and number of coins and started over. During the next 3 minutes he made his second and third tries, each time starting with a five dollar bill. On the second he had 12 bills and coins but an incorrect amount; on the third try, neither was correct. Each time he checked amount and number of coins. During the third 3 minutes he made his fourth try, but made a mistake in adding the amount. Checking through his solution at the beginning of the fourth 3-minute period, he found the error and after 45 seconds offered as a correct solution 1 five dollar bill, 4 one dollar bills, 1 half dollar, 1 dime, 3 nickels, and 2 pennies. Richard made 25 separate moves, but only 6 were needed, as shown in the final solution. Therefore, his efficiency ratio was 6/25, or .24.

In terms of a problem solving sequence set forth by Merrifield, Guil-

ford, Christensen, and Frick (1960), Richard (a) accepted the task as a problem and set a goal of achieving a correct solution, (b) analyzed the available information in relation to his goal, (c) produced a solution and independently rejected it (verification), (d) applied knowledge from this and subsequent solutions to each successive solution produced (reapplication), and (e) finally verified and accepted the correct one.

Michael, one of the more able low IQ boys, had a WISC IQ of 73 and used 805 seconds to "Make 46¢, using 4 coins." Ten coins of each denomination—penny, nickel, dime, and quarter—had been placed on the table before him. (In a pretest it had been established that he had the requisite skill in adding and knowledge of coins to solve the problem.) Michael, as Richard, restated the problem correctly prior to attempting a solution. During the first 3 minutes, Michael took 3 dimes and 10 pennies from the supply, then returned to the supply 3 pennies, one at a time, and offered the remainder as a correct solution. Upon being told that it was incorrect and with all coins returned to the supply, he studied the coins and offered 1 quarter and 2 dimes as a second solution during the second 3 minute period. During the third 3 minutes, he withdrew from the supply 1 quarter, 1 dime, 1 nickel, and 1 penny and offered this as correct. When told it was incorrect and the coins had been returned, he again studied the coins, and during the fourth 3 minutes withdrew 1 quarter, 1 penny, 1 dime, then put the quarter back. During the fifth 3 minutes he again withdrew the quarter and another dime from the supply, thus having 1 quarter, 2 dimes, and 1 penny. This he offered as correct without observably checking ac-

curacy of amount or number of coins. He made a total of 16 coin movements when only 3 were needed; his efficiency ratio was .188. At no point was it necessary to encourage Michael to continue or to keep at the task; however, like the majority of low IQ children, he offered incorrect solutions and did not verify amount of money nor the number of coins.

In what respects are the low IQ most different from the high IQ children? As noted in the above descriptions, the high IQ boy was markedly superior in verification and reapplication, eventually arriving at the correct solution. Further, other low IQ children, different from Michael, did not persist when informed that their solutions were wrong, and after two or three such experiences, manipulated the coins without producing a close approximation of the final correct solution.

### DISCUSSION

If time to reach a correct solution had been used as the criterion of efficiency of the problem solving, no significant difference would have been observed among the three groups. Using the ratio of efficiency as the criterion, a significant difference was observed. The differences of most practical significance among the groups were connected with producing possible solutions, verifying each proposed solution, and applying information from each preceding trial to the subsequent one. As John (1957), Duncan (1959), and others have pointed out, processes such as these used in arriving at solutions are the important variables in understanding efficiency of performance and learning.

Whether more of the low IQ children in this study could learn seemingly simple processes as deciding a

plan for making 13¢ with four coins is not known. Similarly, verifying the amount of money and number of coins does not appear to be difficult, although the high IQ children also tended to do only one or the other on the first try. If these processes can be learned, more specific instruction in school devoted to helping children acquire processes, rather than solutions, would be helpful.

Though not studied intensively, differences within each IQ group were large. For example, the ratio of efficiency varied from .10 to .80 among the high IQ children. Further, 10 high and 16 average IQ children offered incorrect solutions as shown in Table 1. Help given to children of average and high abilities when needed might be especially profitable in achieving improved problem solving performances.

### SUMMARY

Differences among groups of children of low, average, and high IQ were found in observed behaviors as they solved problems, graded in difficulty to their present achievement levels. The high IQ children showed a greater incidence than the average and the low, and the average a greater incidence than the low in noting and correcting mistakes independently, verifying solutions, and using a logical approach; whereas the low IQ showed a greater incidence than the average and the high, and the average a greater incidence than the high in nonpersistence, offering an incorrect solution, and using a random approach. The high IQ children were superior to the low in efficiency of method, as ascertained by the ratio of number of moves made to the least number required for solution of the problem. Differences in performances

among individuals within IQ groups were also large.

#### REFERENCES

- DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397-429.
- JOHN, E. R. Contributions to the study of the problem-solving process. *Psychol. Monogr.*, 1957, **71**(18, Whole No. 447).
- KLAUSMEIER, H. J., CHECK, J., & FELDHUSEN, J. Relationships among physical, mental, achievement, and personality measures in children of low, average, and high intelligence at 125 months of age. *Amer. J. ment. Defic.*, 1960, **65**, 69-78.
- KLAUSMEIER, H. J., & FELDHUSEN, J. Retention in arithmetic among children of low, average, and high intelligence at 117 months of age. *J. educ. Psychol.*, 1959, **50**, 88-92.
- MERRIFIELD, P. R., GUILFORD, J. P., CHRISTENSEN, P. R., & FRICK, J. W. A factor-analytic study of problem-solving abilities. *Rep. Psychol. Lab., U. Southern Calif.*, 1960, No. 22.

(Received October 3, 1960)

# THE RELATION BETWEEN SEVERAL PARENT MEASURES AND THE CHILD'S EARLY ADJUSTMENT TO SCHOOL<sup>1</sup>

GENE R. MEDINNUS

*San Jose State College*

The present paper reports data dealing with first grade readiness and adjustment in first grade. The purpose of this study was to explore the relationship between parental attitudes toward a number of aspects of child rearing, as assessed through questionnaires administered to both parents and through interviews with the mother, and the child's general adjustment in Grade 1.

## METHOD

### Subjects

The subjects in the investigation were a group of first graders and their parents. Two groups of children were identified through the teacher's rating of their adjustment in first grade on a five-point scale, with 5 representing good adjustment, 3 typical of the adjustment of the "average" child, and 1 denoting poor adjustment. The rating was made on the basis of and following the teacher's rating of the children on a 52-item First Grade Adjustment Scale. Those 10 children receiving ratings of 4 and 5 were treated as one group; the other group was comprised of those 9 with ratings of 1 and 2. The mean IQ of the children was 111.6 according to the Stanford-Binet (Form L) intelligence tests administered by the investigator. The socioeconomic status of the families was approximately evenly divided between the upper-lower and lower-middle classes according to Warner's Index of Status Characteristics (Warner, Meeker, & Ells, 1949). At the beginning of the study the mean ages of the mothers and fathers were 32.1 and 34.1 years, respectively, and the corresponding mean years of education were 12.0 and 12.2.

<sup>1</sup>This research was completed under the sponsorship of the Elizabeth McCormick Memorial Fund of Chicago.

### Scales

The two instruments employed to assess parental attitudes and behavior will be described first. Data from these two measures were obtained during the year prior to the child's entrance into first grade.

The parents were asked to complete a form of the Parent Attitude Research Instrument (PARI) developed at the National Institute for Mental Health. The development of the scale, its uses and reliability coefficients have been described in some detail by Schaefer and Bell (1955, 1958). Briefly, the scale is a Likert-type questionnaire consisting of a number of subscales each concerned with various aspects of family life, child rearing attitudes, and relationships. The particular forms of the scale employed consisted of 29 subscales for the mothers and 25 for the fathers. The responses to the items were scored either 4, 3, 2, or 1, depending upon whether the response was "Strongly agree," "Somewhat agree," "Somewhat disagree," or "Strongly disagree," respectively. The questionnaires were left with the parents following an evening home visit by the investigator. They were returned as they were completed.

The Fels Parent Behavior Rating Scales originally were designed in an attempt to identify areas of the home environment which might be expected to bear some causal relation to the development of the child's personality (Baldwin, Kalhorn, & Breese, 1949; Champney, 1941). Lorr and Jenkins (1953) further analyzed the intercorrelations of the Fels scales presented by Roff (1949) in an attempt to determine second-order factors of a broad character which would be useful for conceptualizing the parent-child relationship. Three factors emerged; for convenience in the present discussion these will be labeled as follows: (a) Dependence vs. Independence Encouraging, (b) Democratic vs. Authoritarian, (c) Degree of organization in the home. Those five scales having the highest loadings on each of these three factors were employed in the present study. The ratings were made

by the investigator on the basis of home visits and interviews with the mothers. Interrater reliability was established through previous ratings of 10 homes by the investigator and a graduate assistant.<sup>2</sup> The mean reliability of these ratings for the 15 scales was .80. A list of the 15 rating scales is provided in Table 1.

For the purpose of assessing the subjects' adjustment to the first grade situation, a First Grade Adjustment Scale was developed (Medinnus, in press). Briefly, the following procedure was employed in the construction of the scale. Twenty-five first grade teachers were interviewed concerning factors related to a child's adjustment in Grade 1. The teachers were asked to mention specific traits, abilities, and items of behavior which, in their opinions, characterize or describe good adjustment and poor adjustment in the first grade. The characteristics listed for good and poor adjustment were examined for parallel items which seemed to define opposite ends of a single continuum. Those items for which there was at least a minimum level of consensus were placed on a five-point scale, with 3 representing the midpoint or rating of the "average" child, and 1 and 5 describing poor and good adjustment, respectively. The 52 items were grouped into five major areas: Physical Status, Social Behavior, Emotional Behavior, Intellectual Abilities and Behavior, and Adjustment to Classroom Membership and Requirements. Interrater reliability coefficients for the five sections and for the full scale clustered about .75.

### RESULTS

Those PARI scales which differentiated significantly between the parents of the two groups of children will be mentioned first. The fathers and mothers were compared separately since the subscales and the items were not identical for the two forms. The following scales differentiated between the mothers of the well- and poorly adjusted first graders at the .05 level or beyond: Approval of Activity, Irritability, Dependency of Mother, and Fear of Harming Baby. Mothers of the well adjusted children showed more of the attributes measured by

the first three scales than did the mothers of the poorly adjusted children; the reverse was true for the fourth scale mentioned. No scales reached this significance level for the fathers.

The results for the Fels scales are given in Table 1.

### DISCUSSION

If there is any validity to the premise that the child's behavior and personality are in part causally determined by parental attitudes one would expect wide differences between the attitudes of parents of well- versus poorly adjusted first grade children. That such was not revealed in the present analysis is apparent from the fact that only four scales differentiated between the two groups of mothers and none of the scales reached the significance level for the fathers. One possible interpretation of this finding emphasizes the lack of validity in the parent attitude scale employed. Another interpretation assumes that there is actually no relation between parental attitude and the child's general adjustment in the school situation. The findings with regard to the Fels scales seem to furnish evidence to discount this interpretation.

It is apparent from Table 1 that while only 5 of the 15 Fels scales differentiated between the home environments of the two groups of children, 3 of the 5 were included in the Dependence vs. Independence Encouraging factor. In addition, the parents of the well adjusted group were rated higher on the Solicitousness variable but the difference was not statistically significant. It should be noted also that in each case the homes of the well adjusted children were rated higher in the dependence-encouraging direction. While this would seem to run counter

<sup>2</sup> The writer is grateful to Floyd Ayers for assistance with ratings.

TABLE 1  
RATINGS OF HOMES OF TWO ADJUSTMENT GROUPS

Fels scales	Well Adjusted		Poorly Adjusted		<i>p</i>
	<i>M</i> <sup>a</sup>	$\sigma$	<i>M</i>	$\sigma$	
A. Dependence vs. Independence					
Encouraging					
4.1 General babying	51.3	7.91	42.8	7.81	.05
1.01 Child centeredness of home	55.3	7.85	42.1	6.28	.01
7.1 Solicitousness for welfare	51.0	7.82	45.1	11.69	<i>ns</i>
4.2 General protectiveness	52.7	6.42	40.9	8.48	.01
2.12 Intensity of contact with mother	50.6	8.93	49.0	10.99	<i>ns</i>
B. Autocratic vs. Democratic					
3.15 Nondemocracy of policy	46.4	8.14	52.6	12.87	<i>ns</i>
3.22 Coerciveness of suggestion	47.6	6.92	48.7	12.42	<i>ns</i>
3.14 Nonjustification of policy	44.8	6.22	56.3	10.07	.01
3.11 Restrictiveness of regulations	47.9	8.31	47.3	18.91	<i>ns</i>
8.2 Emotionality toward child	47.5	8.06	53.2	12.93	<i>ns</i>
C. Organization and Effectiveness of Control					
3.12 Readiness of enforcement	48.8	8.15	50.1	13.37	<i>ns</i>
3.13 Severity of actual penalties	47.2	6.57	52.8	12.97	<i>ns</i>
3.3 Accelerational attempt	55.7	7.86	40.7	5.48	.001
1.7 Coordination of household	49.0	7.90	45.4	10.73	<i>ns</i>
1.2 Activeness of home	47.6	8.18	54.9	11.78	<i>ns</i>

<sup>a</sup> The raw score ratings were converted to McCall *T*s.

to the notion of the need for independent behavior in the school setting, the findings obtained here concerning the three scales, General babying, Child centeredness of the home, and Protectiveness, are in accord with the results of other investigations (Baldwin et al., 1949; Bandura & Walters, 1959). In their interpretation of the parent behavior ratings, the Fels researchers identified an "indulgence" pattern consisting of three variables, Babying, Protectiveness, and Solicitousness. High ratings on these variables were interpreted as indicating either one of two processes in the home: "(a) the mother is warmly doting and protective of the child; or (b) she is anxious and restrictive lest the child endanger himself or discommode her" (Baldwin et al., 1949, p. 8). On the basis of the relationship of

these three variables to other rating scales, these investigators concluded that "there is a definite tendency for those parents who are rated high on indulgence to be warm, emotional, and unable or unwilling to give the child either emotional or physical freedom" (p. 8).

While two possibilities were described to explain high ratings on the indulgence variables, the burden of the present discussion is to argue that, in the present study at least, the higher ratings of the homes of the well adjusted first graders on these variables indicate acceptance of the child while the lower ratings of the homes of the poorly adjusted children point to rejection of the child. Great concern with encouraging independence may actually reflect parental rejection of the child. Parental declarations that

the child must learn to "stand on his own two feet" and "fight his own battles" may be attempts to justify nonnurturant and nonsupporting behavior growing out of feelings of rejection.

This interpretation is similar to that of Bandura and Walters (1959) who found that the parents of aggressive boys tended to discourage dependency behavior in their sons and that parental rejection was related to nonnurturant parental behavior. The authors state: "Indeed, parental rejection seems to involve a pervasive form of punishment of the child's attempts to gain emotional support from his parents" (p. 68).

#### SUMMARY

The present paper presented data concerning the relationship between various parental attitudes and behavior and the child's adjustment in first grade.

A comparison was made of the parent attitude scores and Fels home ratings of two groups of children, 10 well adjusted first graders and 9 poorly adjusted first graders.

The following results emerged from the study:

1. Four of the 29 PARI subscales differentiated significantly between the mothers of the groups. None of the scales reached the .05 level for the fathers.

2. Five of the 15 Fels scales differentiated between the home environ-

ments of the two groups of children. Three of the 5 scales were included in the Dependence vs. Independence Encouraging factor.

The results were interpreted as indicating that the lower ratings of the homes of the poorly adjusted children on the several dependence-encouraging variables reflected parental rejection which was a causal factor in the children's poor adjustment to the demands of the first grade situation.

#### REFERENCES

- BALDWIN, A. L., KALHORN, JOAN, & BREESE, FAY H. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, **63**(4, Whole No. 229).
- BANDURA, A., & WALTERS, R. H. *Adolescent aggression*. New York: Ronald, 1959.
- CHAMPNEY, H. The measurement of parent behavior. *Child Developm.*, 1941, **12**, 131-166.
- LORR, M., & JENKINS, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, **17**, 306-308.
- MEDINNUS, G. R. The development of a first grade adjustment scale. *J. exp. Educ.*, in press.
- ROFF, M. A factorial study of the Fels Parent Behavior Scales. *Child Developm.*, 1949, **20**, 29-45.
- SCHAEFER, E. S., & BELL, R. Q. Parental Attitude Research Instrument (PARI): Normative data. Unpublished manuscript, National Institutes of Health, 1955.
- SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Developm.*, 1958, **29**, 339-361.
- WARNER, W. L., MEEKER, MARCHIA, & EELLS, K. *Social class in America*. Chicago: Science Research Associates, 1949.

(Received October 3, 1960)

# ON THE USE OF A SHORT VOCABULARY TEST TO MEASURE GENERAL INTELLIGENCE

JOHN B. MINER

*University of Oregon*

In a prior publication (Miner, 1957) the author summarized the results of some 21 studies which yielded estimates of the relationship between vocabulary tests and measures of what we have come to call *general intelligence*. Most of these studies utilized samples characterized by a rather wide range of ability and employed vocabulary measures of intermediate length such as those contained in the Wechsler or Stanford-Binet tests. The general measures were the 1916 Stanford-Binet, the 1937 Stanford-Binet, the Wechsler-Bellevue scales, the Wechsler Intelligence Scale for Children, the General Classification Tests developed by the Armed Services, and the Otis tests. The median correlation found was .83. On the basis of this evidence, plus the knowledge that the various general measures themselves do not intercorrelate at a much higher level, it was assumed that general intelligence can be adequately predicted from a vocabulary test.

The studies of intelligence differences between various population subgroups and of manpower utilization within the educational and occupational systems of this country which are presented in *Intelligence in the United States* (Miner, 1957) are based on findings obtained with a short vocabulary test. The particular test used was found in three separate investigations (Hagen & Thorndike, 1955; Thorndike, 1942; Thorndike & Gallup, 1944) to have a reliability in the .80 to .85 range. In view of this evidence of reliability and existing knowledge as regards the relationship

between vocabulary and general intelligence, the assumption that an adequate measure of group differences in intelligence was being employed seemed justified.

This conclusion has, however, been criticized on several grounds, perhaps with some justification. The short vocabulary test contained only 20 items whereas the Wechsler and Stanford-Binet measures are considerably longer. Furthermore, the short vocabulary test used a multiple-choice format rather than a free response procedure. A question was raised as to whether the specific vocabulary test used in the studies presented in *Intelligence in the United States* was in fact as valid a measure of intelligence as the presumptive evidence would suggest. No information on the correlation between the short vocabulary test and a more general measure was presented in the book, nor was such information available to the author at the time the book was published.

The present work was carried out with a view to remedying this deficiency and casting further light on the validity of the conclusions reached in *Intelligence in the United States*.

## PROCEDURE AND RESULTS

In connection with a study designed to investigate sales effectiveness, a number of tests were administered to 108 members of the sales organization of a large corporation. The tests employed included Forms A and B of the short vocabulary test originally developed by E. L. Thorndike for the IER Intelligence Scale CAVD, the

Concept Mastery Test (Terman & Oden, 1959) devised to measure the gifted group in adulthood, and the verbal subtests (excluding Vocabulary) of the Wechsler Adult Intelligence Scale. The sample was preselected on intellectual ability since the company characteristically uses tests in the screening of applicants. All of the men had had a high school education at the minimum. Eighty-eight percent had attended college and 68% had graduated. Within the sample the most frequent occupation was that of salesman. However, approximately a quarter of the men were at the management level when tested. None had a Wechsler Verbal IQ below 100.

The intercorrelations are presented in Table 1. Forms A and B of the short vocabulary test have been combined to provide an index of the effectiveness of a longer (40 item) vocabulary measure. Form A was employed in the *Intelligence in the United States* studies. The WAIS Verbal score was computed by combining scaled scores for the Information, Comprehension, Arithmetic, Similarities, and Digit Span subtests.

Table 1 also contains information on the mean scores and standard deviations for each of the tests. The

evidence for selection on intelligence or its correlates is marked. On Form A, for which we have definite information, the mean score for the population as a whole (excluding those aged nine and below) is 10.77. The standard deviation is 3.46.

The difference in means between Forms A and B is reliable ( $t = 5.50$ ,  $p < .01$ ). This finding has been checked on a sample of 64 tabulating machine operators who also took both forms of the test. In this latter group the means were 11.59 and 13.45, respectively, ( $t = 4.95$ ,  $p < .01$ ). Thorndike reports that Form A appears to be somewhat easier at the lower levels and harder at the higher levels than Form B. The results presented here are consistent with such a conclusion.

In view of the known restriction of range in the present sample, it seemed advisable to estimate the size of the correlations that might be expected in the population as a whole. Such figures are more suitable for evaluating the validity of the *Intelligence in the United States* studies which utilized a sample representative of the total population. They are also more suitable for making comparisons with the findings obtained in the studies employing longer vocabulary tests,

TABLE 1  
OBTAINED CORRELATIONS FOR A SELECTED GROUP OF SALES EMPLOYEES  
( $N = 108$ )

Test	MS	SD	Vocabulary Test		Concept Mastery	WAIS Verbal
			Form A	Form B		
Vocabulary Test Forms A & B (40 items)	27.64	3.77	.89	.89	.73	.56
Vocabulary Test Form A (20 items)	13.03	2.11		.59	.64	.47
Vocabulary Test Form B (20 items)	14.61	2.11			.67	.54
Concept Mastery Test	61.50	26.60				
WAIS Verbal	67.22	6.34				.54

which characteristically were based on samples of hospital patients, prison inmates, service inductees, unselected school children, or the adult population.

The estimates in Table 2 which involve the Concept Mastery Test and WAIS were computed using the following formula taken from Gulliksen (1950):

$$R_{XY} = \sqrt{1 - (1 - r_{xy}^2)s_y^2/S_Y^2}$$

In this formula, the standard deviation of the population scores for both forms of the vocabulary test ( $S_y$ ) was taken as 3.46. In all probability the resulting estimates, which average in the high .80s and tend generally to exceed by a small amount the estimate of reliability in the population, are somewhat inflated. Apparently the standard deviations obtained in the selected groups for the two forms of the vocabulary test ( $s_y$ ) are underestimated due to ceiling effects. Inspection of the distributions suggests, however, that the error introduced in this manner is not sizable. It seems almost certain that the population figures would not drop below .75.

The correlation in Table 2 between Form A and Form B is an estimate of reliability in the population as a whole and was derived using Gulliksen's Equation 5 (1950, p. 111). This assumes the two are parallel forms, which is apparently, not strictly speaking, the case. Nevertheless, the figure is consistent with those previously reported. A similar estimate was derived from the tabulating machine operator sample. The corrected reliability coefficient there was .82. The estimates involving Form B are based on the assumption that the standard deviation in the population is the same as for Form A. This assumption in all probability yields excessively conservative estimates of correlation

TABLE 2  
ESTIMATED CORRELATIONS BETWEEN TESTS  
IN THE GENERAL POPULATION

	Vocabulary Test Form B (20 item)	Concept Mastery Test	WAIS Verbal
Vocabulary Test Form A (20 items)	.85	.88	.84
Vocabulary Test Form B (20 items)		.89	.86

size. Lorge (1957) reports a standard deviation of 3.4 for Form A and 4.0 for Form B among United States Army enlisted men.

The above reference also contains information on correlations between the short vocabulary tests and the Army General Classification Test for two representative samples of enlisted men. The samples numbered 868 for Form A and 883 for Form B. The coefficients were .70 and .77, respectively.

#### CONCLUSION

Taken together the Lorge data and the present findings offer rather impressive support to those who have argued for the validity of short vocabulary tests as measures of intelligence. The 20-item tests yield correlations with the more general measures of intelligence which are comparable to those found between the general measures themselves. In the population as a whole the correlation between the short tests and tests of general intelligence appears to be at least .75. This figure compares favorably with that of .83 found for the longer and presumably somewhat more reliable vocabulary tests on which data have been reported previously.

It is hoped that the publication of these findings will result in a greater use of short vocabulary tests in connection with research studies where

practical considerations have heretofore precluded the investigation of relationships between experimental variables and intelligence. Prior assumptions regarding the impossibility of achieving adequate reliability and validity in such short measures were apparently incorrect. There would seem to be little basis for criticizing any study on the score of inadequate measurement techniques, purely because a short verbal measure of intelligence was employed.

#### REFERENCES

GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.

- HAGEN, ELIZABETH P., & THORNDIKE, R. L. Normative test data for adult males obtained by house-to-house testing. *J. educ. Psychol.*, 1955, **46**, 207-216.
- LORGE, I. *Vocabulary Test G-T: Directions and norms*. New York: Columbia University, 1957. (Mimeo)
- MINER, J. B. *Intelligence in the United States*. New York: Springer, 1957.
- TERMAN, L. M., & ODEN, MELITA H. *The gifted group at mid-life*. Stanford: Stanford Univer. Press, 1959.
- THORNDIKE, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26**, 128-135
- THORNDIKE, R. L., & GALLUP, G. H. Verbal intelligence of the American adult. *J. gen. Psychol.*, 1944, **30**, 75-85.

(Received October 12, 1960)

# VISUAL PERCEPTION OF BINARY PATTERNS BY PRESCHOOL CHILDREN AND BY SCHOOL CHILDREN<sup>1</sup>

DOROTHY W. DYER AND E. RAE HARCUM<sup>2</sup>

*College of William and Mary*

This study attempts to relate the appearance of a primacy effect in the perceptual span to progress in learning to read. The primacy effect, as usually defined, is the phenomenon that initial elements of a sequential task produce fewer errors in stimulus reproduction than elements occurring later in the sequence. In the determination of perceptual spans, the superior reproduction of elements at the left within the visual pattern has often been attributed to the development of a learned set or tendency to perceive first those elements at the left within the pattern (Anderson, 1946). Therefore, these elements on the left would be favored by a primacy effect. Since this implied left-to-right viewing sequence is appropriate for reading English, the thesis of the present paper is that it has been developed through practice in reading English. By the same argument, more accurate repro-

duction for stimuli appearing in the right visual field would suggest a right-to-left viewing sequence and, consequently, lesser development in reading skill.

Learning or experience has been found to affect performance on tasks which require the perception of spatial relationships. For example, the tendency to make perceptual reversals decreases as progress is made in learning to read. Smith's (1950) "retarded" readers made approximately 50% more reversal errors than did a group of "advanced" readers. Similarly, inferior and superior readers exhibited differences in the relative number of errors in reproducing tachistoscopic letter-stimuli on the two sides of fixation (Crosland, 1939). Although the superior readers made fewer errors overall, for the stimuli in the right visual field the inferior readers exhibited greater accuracy. Nevertheless, the averages for both superior and inferior readers revealed greater accuracy for elements in the left field of vision. When Anderson (1946) presented nonsense English and Hebrew words to bilingual observers, the observers recognized more English letters to the left of fixation, and more Hebrew letters to the right of fixation. To account for this, Anderson suggested a learned predisposition to perceive first the left-hand letters in English words, and the right-hand letters in Hebrew words—which are, of course, read from right to left.

Heron (1957), also, found fewer errors in recognizing English letters

<sup>1</sup>This article is based on a dissertation submitted by the senior author to the Department of Psychology of the College of William and Mary in partial fulfillment of the requirements for the degree of Master of Arts (Dyer, 1960). This investigation was supported by PHS Research grants (B-1971 and B-2661) from the Institute of Neurological Diseases and Blindness, United States Public Health Service.

<sup>2</sup>Our thanks are due to Anne Smith, Head of the Happy Hours School, Williamsburg, Virginia, and to Jeanne Etheridge, Principal of the Matthew Whaley Elementary School, Williamsburg, Virginia, for their kind cooperation and assistance in the conduct of this research. The authors also wish to thank S. B. Williams and Gertrude Johnson for helpful suggestions and critical reading of the manuscript.

in the left visual field when groups of letters were presented on both sides of fixation. According to Heron, a process involving a spatial-temporal analysis of the stimulus trace resulting from even a single glance favors the letter-stimuli appearing to the left of fixation. This post-exposure process, as Heron calls this sequential analysis of the afterimage, illustrates how sequential viewing habits transferred from the reading situation might affect the perception of tachistoscopically presented material, as well as the perception of material requiring successive eye movements.

Tachistoscopically presented non-alphabetical stimuli produce results similar to those obtained with groups of letters. Harcum (1957), using linear patterns of filled and unfilled circles, found generally superior reproduction for those elements in the left visual field. This result suggests that patterns of meaningless forms may be used to study the development of a so-called normal viewing sequence. The use of such stimuli would allow the investigation of the basic perceptual mechanism without contamination by the meaningfulness, dependencies, or other discriminating characteristics in verbal stimuli.

The present study investigates the development of perceptual organization for visual stimuli using meaningless stimulus-elements simultaneously presented on both sides of fixation. Two hypotheses are tested: observers with no formal reading training reveal little or no difference in reproduction accuracy for stimuli in either visual field; and observers who have had reading training exhibit superior reproduction of stimuli in the left visual field, because they have learned to attend first to stimuli in that hemifield when learning to read.

## METHOD

Sixty-six children with normal vision served as observers. Nursery school ( $N = 13$ ) and kindergarten ( $N = 18$ ) observers were enrolled in the same private school. All first grade ( $N = 18$ ) and second grade ( $N = 17$ ) observers had previously attended that same private school and were pupils at a public elementary school. All children returning a signed parental permission slip were tested, except for the kindergarten children of which only the first 18 were tested.

The apparatus was a Dodge-type tachistoscope. The illumination on the stimulus field and the fixation field, provided by two incandescent light sources, produced a field luminance of about 2 foot-Lamberts. The observers appeared to experience no difficulty in detecting the individual stimulus elements.

Each stimulus template consisted of six horizontally arranged ellipses. The ellipses were .360 inches (31.7') in height, .315 inches (27.8') in width, and were spaced .400 inches (35.3') apart. Stroke width of the ellipses was about .035 inches (3.1'). The fixation cross on the fixation field registered with the center of the template at a distance of 26 inches from the observer's eyes. The target field was 4.5 inches (9.9°) square.

Individual patterns were made by blackening one element of the template on the left and one element on the right of the fixation cross. Each of the six element-positions was filled an equal number of times.

Each observer, viewing binocularly, first reproduced three practice patterns. Each of nine patterns was then observed (in different sequences) with binocular, left-eye, and right-eye viewing. Each observer was assigned one of the six possible viewing orders of the three viewing conditions. An attempt was made to reduce fixational errors by requiring the observer to report the color (which was periodically changed) of the fixation cross prior to exposure of the pattern. Immediately after each .15-second exposure of the pattern the observer pointed out on a blank template before him the elements that corresponded with the ones he had observed as blackened in the stimulus. An error was recorded when an observer incorrectly indicated that an open element was filled, or when he failed to indicate a blackened element.

The intertrial interval was determined by the pace at which the observer chose to

work. Testing time for each observer was usually 15-20 minutes.

A Snellen Chart was used to determine the visual acuity of the observer, and a manoptoscope was used to check his handedness and eyedness. Reading readiness of the kindergarten observers was indicated by performance on Sections 1-4 of the Metropolitan Readiness Tests. Teachers' ratings were obtained to estimate the reading proficiency of the first and second grade observers. Circumstances prevented giving the Metropolitan test to the nursery school children.

### RESULTS

The data from the binocular viewing condition for the four observer groups are shown in Figure 1, in which the means of errors per exposure are plotted as a function of the element-position from the left end of the pattern. The data for the two monocular viewing conditions are not shown because they almost exactly duplicate the binocular data.

As would be expected, there is a decrease in errors associated with the higher educational (maturational) levels.

Since the shapes of the error-curves for the nursery school and kindergarten observers appear basically the same, as do the curves for the first and second grade observers, the observers are categorized for purposes of statistical analysis of the hemifield differences into two general groups of preschool vs. school observers. A small difference in errors favoring the right visual field appears for the preschool observers. When the frequency of observers who exhibit superior performance to the right of fixation are compared to the expected frequency under the null hypothesis, the obtained difference is not significant by the binomial test ( $p > .42$ ). However, the school children exhibit a consistent difference favoring the stimuli in the right visual field. This difference,

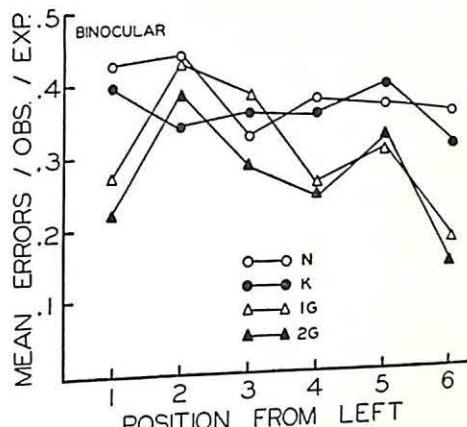


FIG. 1. Means of errors per element-position for each observer group under the binocular viewing condition.

tested as before by the binomial test, is significant ( $p < .012$ ). The marked change in the results of preschool and school children is especially impressive when one considers the wide range of developmental stages represented within each educational level. For example, one child in kindergarten was highly proficient in reading, and his error-curves were quite similar to those of the first and second graders.

The consistent and regular curves in Figure 1 represent the averaging out of individual differences in the relative numbers of errors left and right of fixation. However, these differences in performance between the left and right fields came closer to canceling out for the preschool children than for the school children.

The data of the school children exhibit a decrease in errors at the most eccentric element-positions and for the elements nearest fixation. These curves are similar to those frequently seen in the data of adult observers on a similar task, except for the presence of fewer errors for elements to the right of fixation.

Table 1 presents the comparison of the measures of handedness and eyed-

TABLE 1

RELATION OF EYEDNESS AND HANDEDNESS  
TO TOTAL ERRORS PER OBSERVER AND  
TO PERCENTAGE ERRORS LEFT  
OF FIXATION

Handedness	Measure	Left-eyed	Right-eyed
Left	Total	(N = 8) 49.0	(N = 5) 58.6
	% L	55	53
Right	Total	(N = 20) 52.8	(N = 33) 53.2
	% L	56	53

ness with the total number of errors made by each observer, and the percentage of these errors that occurred for elements to the left of fixation. Neither the total number of errors made by an observer, nor the percentage of his errors which were made for elements left of fixation, are significantly related to laterality preferences.

Table 2 shows a comparison of ratings of higher vs. lower reading proficiency and the occurrence of higher vs. lower relative number of errors to the left of fixation. The frequencies in this two-way classification are based only on the data of the first and second grade observers. These two classifications are not significantly related ( $\chi^2 = 2.40$ ;  $p > .05$ ), although the difference is in the direction of a positive relation between relatively more errors left of fixation and teachers' ratings of above average reading proficiency.

### DISCUSSION

The present study provides evidence in support of the first hypothesis—that the preschool observers do not manifest differences in accuracy of performance between the left and right hemifields. However, the second hypothesis, which specifically predicts the development of superiority for ele-

ments in the left visual field coincident with the onset of instruction in reading, is not confirmed. Fewer errors were made in reproducing stimuli in the right visual field, rather than in the left visual field. Since this significant difference in the relative number of errors appeared rather abruptly with the start of schooling, and since it was not found in the data of either of the two preschool groups, it would seem more reasonably due to instruction in reading rather than to maturation.

Several mechanisms may be proposed for relating instruction in reading to superior reproduction for stimuli in the right visual field. Children who have had instruction in reading may tend to look toward the right end of the visual pattern (i.e., fixate incorrectly) in this task. But this does not seem likely. Similarly, the children may have acquired a right-to-left viewing sequence in the early stages of learning to read, but this also seems unlikely. On the other hand, the children may have developed a left-to-right scanning sequence as predicted, but, for observers of this developmental level, recency effects may be stronger than primacy effects. This latter mechanism seems to be the most plausible in view of the small, although not statistically significant, positive relationship between ratings of reading proficiency and greater

TABLE 2  
RELATION OF READING RATINGS TO THE  
PERCENTAGE ERRORS TO THE LEFT OF  
FIXATION FOR FIRST AND SECOND  
GRADE OBSERVERS

% Left	Average reader or below	Above average reader
Above 50%		
50% or below	9 6	13 7

numbers of errors to the left of fixation.

### SUMMARY

Very simple binary patterns were observed by nursery school, kindergarten, first grade, and second grade school children with left-eye, right-eye, and binocular viewing. It was hypothesized, on the basis of data previously obtained from adult observers, that the relative accuracy for reproducing elements to the left of fixation would increase with school training in the left-to-right reading sequence of English.

The preschool observers exhibited nonsignificantly greater accuracy for stimuli in the right hemifield. Contrary to the hypothesis, significant right hemifield superiority was observed in the school children. Apparently, perceptual mechanisms of children at the age at which they start school do not wholly duplicate adult mechanisms.

No important effect of viewing-eye condition, or of eyedness, handedness,

or "conflicting" eyedness-handedness of the observers was found.

### REFERENCES

- ANDERSON, I. H. The effect of letter-position on range of apprehension scores, with special reference to reading disability. *U. Mich. Sch. Educ. Bull.*, 1946, 18, 37-40.
- CROSLAND, H. R. Superior elementary-school readers contrasted with inferior readers in letter-position, "range of attention," scores. *J. educ. Res.*, 1939, 32, 410-427.
- DYER, DOROTHY W. Perception of binary visual patterns by pre-school children and by school children. Unpublished master's dissertation, College of William and Mary, 1960.
- HARCUM, E. R. Three inferred factors in the visual recognition of binary targets. In J. W. Wulfeck & J. H. Taylor (Eds.), *Form discrimination as related to military problems*. (NAS-NRC Publ. 561) Washington, D. C.: NAS-NRC, 1957. Pp. 32-37.
- HERON, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, 70, 38-48.
- SMITH, LINDA C. A study of laterality characteristics of retarded readers and reading achievers. *J. exp. Educ.*, 1950, 18, 321-329.

(Received November 4, 1960)

## FIXED SEQUENCE VERSUS BRANCHING AUTOINSTRUCTIONAL METHODS

HARRY F. SILBERMAN, RALPH J. MELARAGNO, JOHN E. COULSON,  
AND DONALD ESTAVAN

*System Development Corporation*

Two basic autoinstructional methods may be identified. The one most frequently used displays a fixed sequence of teaching items, requiring all subjects to traverse all items (Skinner, 1958). An alternative method has been suggested (Coulson & Silberman, 1960; Crowder, 1960) which uses a variable sequence of items, allowing each student to branch to remedial materials when needed.

With a fixed-sequence teaching machine, adaptation to individual differences is provided only by the variable rate at which students move through the sequence. Variable-sequence teaching machines provide for individual differences by using a variety of criteria, such as error rate and error type, for branching to additional remedial items and for skipping over redundant items. Each student generates his own unique sequence of items to effect greater learning efficiency. The basis for the design of an autoinstructional method in which the sequence of instructional material is tailored to the needs of each student is the notion that there is no single best item sequence nor one best method of presenting the materials. It is assumed that different students require different items with varying amounts of review at different times during training.

Coulson and Silberman (1960) attempted to evaluate the effectiveness of branching versus fixed sequence instructional methods. Using a manually operated machine they compared a fixed item sequence with a variable

sequence in which some items were skipped by students who made no errors. The findings showed that students trained with the very simple branching procedure used in that study learned as well as students trained without branching, and also learned in a shorter time. The amount of instructional material was not controlled, since subjects in the branching group had skipped certain items seen by the fixed-sequence group. In such a study, obtained differences can be attributed to the difference in items presented as well as to the effects of adjusting the instructional sequences to the needs of individual students.

In the Coulson and Silberman (1960) study, the wishes of the student were not considered in determining whether he was branched back to preceding items or skipped ahead to subsequent material. Nor was error rate over a sequence of items used as a branching criterion. The branching depended entirely on the errors made on individual items.

The present study was concerned with two methods of branching. In Experiment I, a procedure that permitted the student to branch at his own option was evaluated, while in Experiment II the use of errors as a branching criterion was attempted. The hypothesis tested in Experiment I was that optional-branching sequences would be superior to a fixed sequence; the hypothesis in Experiment II was that an error-generated branching sequence would be superior

to a fixed sequence when identical items are used in each sequence.

### EXPERIMENT I

Two hypotheses relating to optional branching were tested. The first was that a fixed-sequence program would be less effective than the same items cast as statements in textbook format through which the student could branch at his own option. The ease of reviewing and the facility for skipping ahead with the textbook format were the bases of this hypothesis. The second hypothesis was that performance on a program which allowed the student to review previous material, one item at a time, would also exceed performance on a fixed-sequence program.

#### Procedure

Students from five high schools in Los Angeles County served as paid subjects. Subjects were obtained from junior and senior classes, and were pretested with the Henmon-Nelson Test of Mental Ability (Form A, Grades 9-12). Random assignment to three treatment groups resulted in 7 males and 10 females being assigned to two groups, and 6 males and 11 females being assigned to the third group, providing 17 subjects in each group.

Sixty-one multiple-choice teaching items on some topics in logic were prepared and given a preliminary trial in a number of secondary schools. Items were revised, typed on  $5 \times 8$  cards, and mounted in loose-leaf binders for the experiment. Topics covered included simple and compound statements, connectives and their combinations, and arguments.

Three instructional methods were used. The first group, designated "fixed sequence," (Henmon-Nelson Mean = 114.1) received the 61 teaching items in a fixed sequence. Subjects in this group were instructed to give a covert response to each item in turn, and to compare their answer with the correct answer, which was inscribed on the back of the card. Subjects were further instructed not to turn back to preceding items. The second group, designated "back branching," (Henmon-Nelson Mean = 113.1) received the same 61 teaching items but was permitted to review preceding items by

backing up one item-card at a time. Subjects in this group were encouraged to review earlier material and were instructed not to skip any items, so that they would view each item at least once. The third group, designated "textbook," (Henmon-Nelson Mean = 111.3) received a type-written text produced by rewriting the original items. Each item was converted from a question into a statement by incorporating the correct answers, eliminating distractors and filling in blanks. These statements were then organized into paragraphs. Subjects in the third group were instructed to study the material in whatever manner they chose. All subjects were observed through one-way windows to insure that they were following directions and, additionally, to check on whether subjects in the back-branching group exercised the branching capability.

Mastery of the topics covered in the teaching session was tested with 24 free-response and 24 multiple-choice test questions. Half of the questions were similar to the actual training materials and half consisted of situations which required application or transfer of learning. Split-half reliability of the posttest was .93 (after Spearman-Brown correction). The test was administered immediately after the training session.

#### Results

Comparison of the criterion scores (Table 1) for the textbook and the fixed-sequence groups confirms the hypothesis that the textbook method, which permits the student to branch at his own option, is superior. The second hypothesis, that the back-branching method would be superior to the fixed-sequence method, was not confirmed.

TABLE 1  
POSTTEST SCORES AND TRAINING TIME

Treatment	N	Criterion Posttest		Time (in minutes)	
		M	SD	M	SD
Fixed Sequence	17	28.4	8.7	32.8	10.7
Back Branching	17	30.8	9.7	31.8	9.6
Textbook	17	33.7	6.4	31.5	11.1

TABLE 2  
CORRELATIONS OF APTITUDE, TRAINING  
TIME, AND POSTTEST SCORE

	Fixed Sequence	Back Branching	Textbook
Aptitude vs. Score	.45	.39	.67**
Aptitude vs. Time	-.52*	-.45	-.31
Time vs. Score	-.14	.03	-.04

\*  $p < .05$ .

\*\*  $p < .01$ .

For the first hypothesis, that the textbook method would be superior to the fixed-sequence method,  $t = 1.79$ ,  $p < .05$ . For the second hypothesis, that the back-branching method would be superior to the fixed-sequence method,  $t = .80$ ,  $p > .10$ .

Correlations of Henmon-Nelson aptitude scores, criterion posttest scores, and training times are given in Table 2.

## EXPERIMENT II

A branching program, using error rate as a branching decision criterion was compared with a fixed-sequence program with instructional materials equated for the two groups. It was hypothesized that performance of subjects receiving a branching sequence would be superior, with respect to scores on a criterion posttest, to that of students receiving a fixed sequence. This was predicted on the basis of the teaching machine's responsiveness to the student's performance. By providing extra remedial items to students who have difficulty, the machine tends to minimize the number of items seen by the brighter student and to minimize the number of errors on training items made by the low aptitude student. The overall effect is intended to increase the amount of learning in a given time for each individual.

## Method

**Apparatus.** A high speed, general purpose digital computer (Bendix G-15), equipped with a 600-slide, random-access slide projector, and an electric typewriter, was used as a teaching machine. Each slide contained one multiple-choice instructional item. Subject used the typewriter to insert his answer and receive knowledge of results.

The computer was programmed to select a variable sequence of slides from a projector in response to different errors made by subject. If subject was exceeding the number of errors that had been prespecified for a particular topic, he would be branched to a set of less difficult items. If he did poorly on this set he would be shunted to still easier sets on the same topic before moving on to the next topic. Some of the multiple-choice items contained distractors the choice of which indicated a particular misunderstanding. Subjects who selected these options were branched immediately to remedial slides designed to clear up that point.

**Instructional Materials.** Four hundred and eleven multiple-choice items on logic (Silberman, Coulson, Melaragno, Estavan, & Englund, 1960) were prepared and given a preliminary trial in a number of secondary schools. Items were revised and put onto 35-mm. slides for use in the random-access projector.

Topics covered by these items included simple and compound statements, logical connectives, and six argument forms. Instructional items were arranged into difficulty levels. Items at one level were written so as to contain a maximum amount of information, and were the most difficult items. The items at the second level were essentially in the same format as those of the first level except that pertinent information was generally distributed over more slides than at the first level. The same was true of the third and fourth levels; here the items were designed to be relatively easy to answer, in that the step between items was kept small and the prompts were increased.

**Subjects.** Thirty-six students (21 males and 15 females) were selected from four high schools in Los Angeles County. Subjects had either just graduated or were in their last year of high school. Subjects were pretested with the Henmon-Nelson Test of Mental Ability (Form A, Grades 9-12) and were randomly assigned to two groups. Subjects were naive with respect to the subject matter (logic) used in the lesson.

**Procedure.** One treatment group was des-

gnated the branching group and the other the fixed-sequence or nonbranching group. Members of the branching group received sequences of items determined by the errors that were made during the teaching session. The computer kept cumulative error records and also a record of errors on individual items. Throughout the session the machine selected an appropriate sequence of instructional items for each student based upon error criteria. In this way, the bright student was given a minimum of material, while the slower student was given extra, remedial material on topics in which errors were committed.

Each member of the fixed-sequence group was paired at random with one member of the branching group. The unique sequence of items which each subject in the branching group generated was presented to his mate in the fixed-sequence group. Thus, pairs of subjects in the fixed-sequence and the branching groups received *identical* item sequences. The difference between the two groups was that the machine was responsive to the errors made by subjects in the branching group but was not responsive to errors made by subjects in the nonbranching group.

If the brighter member of a pair of subjects was in the branching group, the resulting sequence was expected to be too brief to provide effective instruction for his mate in the fixed-sequence group. If the brighter member of a pair of subjects was in the fixed-sequence group, the sequence which he received was expected to be longer than necessary for him, resulting in wasted time and boredom.

Knowledge of results, which consisted of a typewriter printout of the number of the correct answer, was the same for both groups. One subject was trained at a time.

Subjects were given an explanation of the operation of the equipment and permitted to familiarize themselves with the machine on the first two introductory items. In the middle of the session, subjects were given a short break. Immediately after training was completed, the criterion posttest was administered to each subject. No time restrictions were placed on subjects either during the training period or during the posttest.

*Criterion Test.* Mastery of the topics covered by the training items was tested with 51 multiple-choice test items and 44 free-response test items. Half of the test items were quite similar to the actual training materials and the other half required application or transfer of learning to situations not encountered in the training series.

TABLE 3  
APTITUDE, TRAINING, AND POSTTRAINING PERFORMANCE

		Treatment Groups <sup>a</sup>	
		Branching	Fixed Sequence
Number of Instructional Items	<i>M</i>	156.00	156.00
	Range	69-248	69-248
Training Errors	<i>M</i>	24.66	21.94
	<i>SD</i>	17.30	13.57
Training Time (minutes)	<i>M</i>	78.83	81.61
	<i>SD</i>	27.34	29.21
Henmon-Nelson IQ's	<i>M</i>	123.61	122.61
	<i>SD</i>	19.16	17.89
Posttraining Criterion Scores	<i>M</i>	66.33	64.33
	<i>SD</i>	15.93	16.42

<sup>a</sup> *N* = 18 for each treatment group.

The split-half reliability of the test was .97 (after Spearman-Brown correction).

### Results

Table 3 gives data on the number of items traversed by subjects, the number of errors made on instructional items, and the training time. The mean time required for each item was 17 seconds. Table 3 also contains information on the Henmon-Nelson aptitude pretest and the criterion posttest.

Covariance analysis of criterion posttest scores, controlling training time and aptitude, yielded no significant difference between the branching and the fixed-sequence groups.

Intercorrelations of aptitude, criterion score, training time, and instructional errors are given in Table 4.

### DISCUSSION

While both experiments were concerned with branching, they are not

TABLE 4

## INTERCORRELATIONS OF APTITUDE, CRITERION SCORES, TRAINING TIME, AND INSTRUCTIONAL ERRORS

	Branching	Fixed Sequence
Aptitude vs. Score	.763*	.845*
Aptitude vs. Errors	-.677*	-.769*
Aptitude vs. Time	-.833*	-.407
Time vs. Score	-.682*	-.179
Errors vs. Score	-.904*	-.846*
Errors vs. Time	.793*	.472

\*  $p < .01$ .

directly comparable in that different apparatus and procedures were used.

Overall, the results of Experiment I indicate that the relationship between flexibility of materials and rate of learning favors the flexible materials over the fixed sequence. Certain implications for further exploration are evident when this study is compared with those by Holland (1960) and Cook and Spitzer (1960), which contain features similar to it.

With a semester course in introductory psychology as his subject matter, Holland also used three instructional groups. Members of the first group were trained in the conventional manner, with confirmation of the correct answer following subject's overt response. In his second group, subjects received no confirmation following their response.

Holland found no difference in performance between the confirmation program and the same program without the confirmation. Both of these programs were superior to a no-response, control program which only required subjects to read complete statements. Holland interpreted his results as lending support for the principle that an overt response has to occur for efficient learning. Insofar as Holland's control group can be thought of as using a prompting procedure, his

results are in support of confirmation as a better teaching method.

Cook and Spitzer (1960) reported results of a study using a short list of paired-associates which seem to conflict with Holland's findings. A  $2 \times 2$  design was employed in which the experimental conditions were prompting and confirmation, with overt response or no response used in each condition. Cook and Spitzer concluded that overt responses interfered with learning and that the prompting procedure was superior to confirmation because of the enforced delay between stimulus and response terms in the confirmation method.

The superiority of the textbook group over the fixed-sequence group in Experiment I lends support to the prompting side of the prompting-Confirming controversy, when no overt response is required, since the textbook condition corresponds to the prompting procedure while the fixed-sequence condition used a confirmation procedure.

Since the three studies involved materials which were dissimilar in length and quality, it remains for further research to compare the influences of response requirement, S-R delay, and optional selection of material.

Two possible conclusions can be obtained from Experiment II. The first is that a teaching machine which branches on the basis of errors is not significantly superior to a machine which displays a fixed sequence of items. An alternative explanation is that the error criteria used for branching in this experiment, and the remedial items to which a student who was having difficulty was branched, were inefficient and acted to obscure any real differences between the two methods. Certain of the experimental results lend support to the second conclusion.

Analysis of the criterion tests shows that low aptitude subjects in both groups failed to learn much of the material. While this had been predicted for the fixed-sequence group, it also had been expected that members of the branching group, who would supposedly receive a sequence uniquely tailored to their individual needs, would receive the training necessary to insure their learning the material. This was the case for only two (of seven) subjects with low aptitudes in the branching group. Both of these subjects were frequently branched to remedial items, and both scored above their group's mean on the criterion test. The five remaining low aptitude subjects in the branching group failed to score at or near their group's mean. All six subjects in the fixed-sequence group with low aptitudes fell well below their group's mean.

Three factors which may account for the ineffectiveness of the remedial items for low aptitude subjects are: the limited variety of these items, the inability of subjects to branch out of the remedial series when their difficulties had been resolved, and the inappropriateness of the particular measure of individual differences (errors) used as the criterion for branching decisions.

Since the instructional items in the remedial branches were very similar to those in the main series, except that the subject matter had been divided into small steps, the mode of instruction in Experiment II was of limited versatility. That is, no attempt was made to alter the teaching method for a given subject who was having difficulty. It is postulated that the low aptitude subjects who scored well actually received remedial material which assisted them, while the low aptitude subjects who did poorly were not assisted because the machine sim-

ply was not flexible enough to find a set of items adequate for them.

When a low aptitude subject was branched to remedial material, no attempt was made to evaluate the subject's correct responses in order to branch him back to the main set of items when he demonstrated proficiency. Instead, subject was obliged to remain in the remedial branch until he had negotiated all of the items in that branch. For many of the low aptitude subjects, this meant that they were forced to see more items than necessary.

In Experiment II it was assumed that the dimensions of individual differences measured by errors on instructional items interacted with the effectiveness of the item sequence presented. The machine was, consequently, programmed to treat subjects differentially (i.e., giving different item sequences) with respect to the dimension of error rate. The results show that the particular method of providing for individual differences used in this experiment was not a sufficient condition for effective learning of the logic lesson. Perhaps the common principle of teaching that one should "provide for individual differences" needs to be qualified with the specific conditions for its accomplishment. It may be conjectured that some measures such as response latency or subject's self-evaluation are more appropriate than error rate, and that the computer should have considered these behavior measures for its branching decisions instead of, or in addition to, errors.

#### SUMMARY

The experiments investigated the effectiveness of two methods of branching, as measured by posttraining criterion tests. In Experiment I, three groups of 17 subjects were used to test

two hypotheses concerning optional branching. A fixed-sequence group received items in fixed order; a back-branching group receiving the same items as the first group, was permitted to back up one item at a time to review earlier items; a third group received the same items cast in statement form and organized into paragraphs permitting subjects to choose material at their own option. A significant difference on a posttest in favor of the third group was obtained when the first and third groups were compared.

In Experiment II, a computer-controlled teaching machine was used to evaluate the effectiveness of adapting sequences of teaching items on logic.

Members of a branching group received sequences of items determined by the errors that were made during instruction. Each member of a fixed sequence group was paired at random with one member of the branching group. The particular sequence of items which each subject in the branching group received was presented to his mate in the fixed-sequence group. Covariance analysis of

criterion scores using aptitude and training time as control variables yielded no significant difference between the branching and fixed-sequence conditions. A discussion of the limitations of the study and of possibilities for further research on this problem was presented.

#### REFERENCES

- Cook, J. O., & Spitzer, M. E. Supplementary report: Prompting versus confirmation in paired associate learning. *J. exp. Psychol.*, 1960, **59**, 275-276.
- Coulson, J. E., & Silberman, H. F. Effects of three variables in a teaching machine. *J. educ. Psychol.*, 1960, **51**, 135-143.
- Crowder, N. A. Automatic tutoring by intrinsic programming. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington: National Education Association, 1960.
- Holland, J. G. Program design and use. Paper read at American Psychological Association, Chicago, September 1960.
- Silberman, H. F., Coulson, J. E., Melaragno, R. J., Estavan, D. P., & Englund, D. E. *Teaching machine program: Logic*. (TM-534) Santa Monica, Calif.: System Development Corporation, 1960.
- Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969-977.

(Received November 7, 1960)

## A RE-EXAMINATION OF COLLEGE PRODUCTIVITY<sup>1</sup>

ALEXANDER W. ASTIN<sup>2</sup>

*National Merit Scholarship Corporation*

Several recent investigations have attempted to evaluate what part the undergraduate's educational experience plays in stimulating him to go on for the PhD degree. The first of these, the studies of Knapp and Goodrich (1952) and Knapp and Greenbaum (1953), defined the productivity of any undergraduate college as the percentage of its graduates who eventually obtain the PhD. In a subsequent study, Holland (1957) showed that an institution's productivity is related in part to the academic ability of the students who enroll; that is, the highly productive institution has a greater proportion of talented students among its undergraduates than does the less productive institution. Following this lead, Thistlethwaite (1959a, 1959b) equated colleges with respect to student ability by adjusting each college's PhD productivity rates in terms of its "talent supply" (that is, the estimated academic ability of the student body). He found correlations as high as .83 between these adjusted productivity rates and certain "environmental press" characteristics of 36 colleges. Productivity of PhD's in the natural sciences (NS) was correlated with a somewhat different set of press than was productivity in the arts, humanities, and social sciences (AHSS).

The present study attempts to extend this work by exploring the following questions:

<sup>1</sup>This study is a part of the research program at the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation.

<sup>2</sup>The author is indebted to John L. Thistlethwaite, Laura Kent, and Donald L. Thistlethwaite for reviewing the manuscript.

1. Is ability the only quality of the student body which is related to a college's productivity? Are there not other systematic differences among the student bodies enrolling at various institutions which are related to PhD productivity? One such difference might be in the proportion of entering students who plan careers in either NS or AHSS. Certainly any college which is attended predominantly by students who plan careers in the natural sciences can be expected to have a relatively high proportion of graduates who go on for the PhD in NS, independent of any special effects of the college environment. Conversely, this same college would probably have a low productivity rate of AHSS PhD's. Two other characteristics of entering students which might affect the productivity of a given school are the ratio of men to women, and the proportion of students who already intend to get the PhD degree.

2. If the educational goals, major fields, and sex ratios of entering students are related to productivity, are they also responsible in some measure for the correlations found earlier between "college press" and productivity? These correlations have been interpreted as evidence for a causative role of the college on the student's motivation; e.g., "the results... indicate that the teacher plays an important role in motivating talented undergraduates to seek advanced degrees" (Thistlethwaite, 1959a, p. 75); "The analysis confirms once again the view that one type of college press stimulates achievement in the natural sciences, while a different type facilitates achievement in the arts,

humanities, and social sciences" (Thistlethwaite, 1959b, p. 188). However, since the information on college press came from students at each institution, it may reflect differences among student bodies which existed prior to the students' entrance into college, differences which are not a product of any unique qualities of the college environment. Before we can accurately assess the motivating function of the college itself, we must find out the extent to which these "pre-college" characteristics of the student body are related to the college's productivity.

Specifically, this study examined:

1. The extent to which productivity rates of NS and AHSS PhD's (adjusted for talent supply) are a function of: (a) the proportion of students who, prior to entering college, intend to major in natural science; (b) the proportion of students who, prior to entering college, aspire to the PhD degree; and (c) the ratio of men to women among students enrolling in the college.

2. The extent to which relationships previously obtained between adjusted PhD productivity and college press are modified, when the effects of these additional precollege variables are removed.

#### METHOD

##### *College Input Measures*

Information on three precollege or input variables—the student's sex, probable college major, and highest degree aspired to—was drawn from a questionnaire which had been mailed to a 10% random sample of all high school students who scored above the 64th percentile on the 1959 National Merit Scholarship Qualifying Test. Of the students who returned the questionnaires, 2,884 attended one of the 36 institutions about which college press data had been obtained in the earlier studies. The number of students attending each institution ranged from 15 to 208, with a median of 72.

The following three input measures were obtained for each of the 36 colleges: the percentage of students who originally planned to major in natural science (% NS Majors), the percentage of students aspiring to the PhD degree (% PhD Aspirants), and the percentage of males among the entering students (% Males).<sup>3</sup>

#### RESULTS

Table 1 shows the correlations between the three input measures and the adjusted productivity indexes of the 36 colleges.<sup>4</sup> It is evident that a significant amount of the variation in adjusted PhD productivity can still be accounted for on the basis of differences among the entering students. To obtain a productivity index which is independent of the quality of the student body, then, requires the use of corrections for variables other than just aptitude level.

The intercorrelations of the three input variables are given in Table 2. To estimate the unique contributions of these variables to productivity, their partial correlations with productivity rates were computed (Table 3).

<sup>3</sup>In order to make these distributions more nearly normal, all percentages were converted by means of the arcsin transformation (Guilford, 1954, pp. 574-576). Since these input measures were based on limited samples of the 1959 freshman classes at the 36 colleges, two checks were made on the representativeness of the NMSQT samples. First, % Males was correlated with the percentage of males among the entire freshman class entering each of these 36 institutions in Fall, 1958 (as reported by United States Office of Education, 1959). This coefficient was .97. Second, mean Scholastic Aptitude Test scores, which were available on the entire 1956 freshman classes enrolling at 22 of the 36 colleges, were correlated with the corresponding mean NMSQT scores of the present samples. This coefficient was .86.

<sup>4</sup>In computing these correlations it is necessary to assume that the relative status of the 36 colleges on the three input variables has not changed greatly over a 10- to 15-year period. In earlier reports similar assumptions were made with respect to talent supply and "college press."

Table 3 makes it clear that % Males relates to neither NS nor AHSS productivity, beyond the combined contribution of % NS Majors and % PhD Aspirants. (This would suggest, in addition, that differences in the proportions of men and women who get the PhD can be accounted for by differences in their precollege levels of aspiration and intended fields of study.) These findings prompted the decision to omit % Males as an input variable, and to deal only with % NS Majors and % PhD Aspirants.

Table 4 shows the correlations of % NS Majors and % PhD Aspirants with those Faculty and Student Press Scales which were previously reported as significantly related to adjusted

TABLE 1  
CORRELATIONS BETWEEN THREE INPUT  
VARIABLES AND THE ADJUSTED<sup>a</sup> PhD  
PRODUCTIVITY RATES OF 36  
INSTITUTIONS

	Input Variables		
	% NS Majors	% PhD Aspirants	% Males
Adjusted Productivity in Natural Science (NS)	.60**	.68**	.45**
Adjusted Productivity in Arts, Humanities, and Social Science (AHSS)	-.69**	.12	-.20

<sup>a</sup> "Adjusted" productivity is defined as the discrepancy between a college's expected rate of producing PhD's (as predicted from its talent supply) and its actual productivity rate (Thistlethwaite, 1959b).

\*\*  $p < .01$ .

TABLE 2  
INTERCORRELATIONS OF THREE  
INPUT VARIABLES

	% PhD Aspirants	% Males
% NS Majors	.46**	.50**
% PhD Aspirants		.49**

\*\*  $p < .01$ .

TABLE 3  
SECOND-ORDER PARTIAL CORRELATIONS  
SHOWING THE INDEPENDENT CONTRI-  
BUTION OF EACH INPUT VARIABLE  
TO ADJUSTED PhD PRODUCTIVITY

	Input Variables		
	% NS Majors	% PhD Aspirants	% Males
Adjusted NS Productivity	.41*	.54**	.03
Adjusted AHSS Productivity	-.83**	.66**	.00

\*  $p < .05$ .

\*\*  $p < .01$ .

TABLE 4  
CORRELATIONS OF TWO INPUT VARIABLES  
WITH 13 COLLEGE PRESS SCALES AT  
36 INSTITUTIONS

	Input Variable	
	% NS Majors	% PhD Aspirants
Student Press		
Humanism	-.39*	.36*
Breadth of interests	-.47**	.08
Reflectiveness	-.29	.45**
Participation	.12	-.41*
Aggression	.59**	.13
Scientism	.48**	.64**
Social conformity	-.36*	-.52**
Faculty Press		
Excellence of social science faculty and resources	-.71**	-.04
Flexibility of curriculum	-.56**	.03
Energy and controversiality of instruction	-.27	.31
Informality and warmth of student-faculty contacts	-.04	.55**
Closeness of supervision	.03	-.45**
Effectiveness of teaching methods	-.41*	-.34*

\*  $p < .05$ .

\*\*  $p < .01$ .

productivity rates (Thistlethwaite, 1959b, pp. 187-188). Nine of the 13 college press scales are significantly ( $p < .01$ ) related to one or both of the

TABLE 5  
CONTRIBUTION OF TWO INPUT VARIABLES TO CORRELATIONS BETWEEN COLLEGE  
PRESS AND ADJUSTED PHD PRODUCTIVITY

	Adjusted NS Productivity		Adjusted AHSS Productivity	
	Original Correlation <sup>a</sup>	Partial Correlation <sup>b</sup>	Original Correlation <sup>a</sup>	Partial Correlation <sup>b</sup>
Student Press				
Humanism	-.07	-.25	.71**	.39*
Breadth of interests	-.27	-.26	.67**	.42*
Reflectiveness	-.05	-.38*	.67**	.40*
Participation	.00	.21	-.53**	-.42*
Aggression	.43**	.29	-.54**	-.13
Scientism	.59**	.18	.03	.33
Social conformity	-.42*	-.10	.03	-.03
Faculty Press				
Excellence of social science faculty and resources	-.42*	-.33	.83**	.55**
Flexibility of curriculum	-.31	-.24	.68**	.33
Energy and controversiality of instruction	-.13	-.35*	.58**	.35*
Informality and warmth of student-faculty contacts	.43**	.32	.40*	.14
Closeness of supervision	-.22 <sup>c</sup>	.00	-.38*	-.23
Directiveness of teaching methods	-.42**	-.16	.24	.12

<sup>a</sup> From Thistlethwaite (1959b, pp. 187-188).

<sup>b</sup> Independent of % NS Majors and % PhD Aspirants.

<sup>c</sup> These correlations were inadvertently interchanged in the original report.

\* $p < .05$ .

\*\* $p < .01$ .

input variables, and three of the other four press scales have borderline ( $p < .05$ ) relationships with input.

In order to remove the effects of these two input variables from the correlations previously obtained between college presses and PhD productivity, second-order partial correlations were computed. Table 5 shows these correlations, along with the original correlations given in the earlier report.<sup>5</sup> It can be seen that all the larger correlations tended to shrink when the two input variables were partialled out. Of the 11 previously significant ( $p < .01$ ) correla-

<sup>5</sup> Second-order partial correlations in a relatively small sample such as this should be regarded as very rough approximations. They are presented here only to show trends in the effects of student input on the original correlations.

tions, only one (between AHSS Productivity and "Excellence of social science faculty and resources") remained significant; five of these, all involving AHSS Productivity, were reduced to borderline significance ( $p > .01, < .05$ ), and five to nonsignificance. In addition, all five previously borderline correlations were reduced to nonsignificance. The only reversals of the trend toward shrinkage were two previously nonsignificant correlations with NS Productivity, which became borderline as a result of the partialing procedure.

#### DISCUSSION

These findings seem to warrant two conclusions. First of all, adjusting the productivity rates of undergraduate institutions solely in terms of differ-

ences in the aptitude levels of their students does not result in productivity indexes which are independent of student quality. At least two other characteristics of entering student bodies—the percentage planning to major in NS and the percentage aspiring to the PhD degree—are significantly related to adjusted productivity rates. Secondly, these input variables account for a substantial share of the relationships previously found between college environments and adjusted productivity. In short, much of the variation in PhD productivity which had been previously attributed to the motivating effects of the colleges now appears to be a function of differences among the student bodies which enroll.

An additional question posed by these analyses concerned the justification for using two productivity indexes. The high correlations of % NS Majors with both productivity rates (Table 1) suggested that breaking down productivity in this way introduces extraneous variability due to differences in the students' fields of interest. To check this notion more directly, the difference between each school's productivity rates was obtained (that is, NS rate - AHSS rate). The correlation between this difference score and % NS Majors was .84. In light of this finding, the total (unadjusted) PhD productivity rate (see Thistlethwaite, 1959b, p. 184) was computed for each college. The correlation between total productivity and % PhD Aspirants was .81. (% Males and % NS Majors were not significantly related to total productivity.) It seems apparent, then, that about two-thirds of the variation among these 36 colleges in total PhD productivity can be accounted for on the basis of a single input variable.

These findings have several implications. For one thing, they highlight

one of the difficulties in attempting to assess the influence of the college on its students: differential "effects" of colleges may well turn out to be spurious, unless the relevant student input variables are controlled. For another, these results suggest that it may be necessary to reinterpret some of the results of other studies, not only more recent ones of college "press" (Thistlethwaite, 1960), but also those which utilize adjusted college productivity rates (e.g., Holland, 1960).

From the available evidence, it is not clear whether or to what extent the undergraduate institution stimulates or inhibits PhD productivity. The high correlations found between input and productivity indicate that the college plays a relatively smaller role in this sample than was previously supposed; however, we would have to study larger samples of students and colleges before we could determine precisely the relative contributions of input variables and college environments. Moreover, since this sample of colleges is far from representative,<sup>6</sup> due caution should be used in attempting to generalize the results to other institutions.

#### SUMMARY

The PhD-producing rates of 36 colleges were subjected to further study. It was shown that a college's productivity rates are related to characteristics of its entering students other than just their academic ability; namely, the percentage planning to major in natural science and the percentage aspiring to the PhD degree. When the effects of these two input variables are partialled out, correlations previously obtained between college press and productivity rates are reduced considerably in size.

<sup>6</sup> These 36 colleges were selected initially because they were attended by relatively large numbers of National Merit Scholars.

## REFERENCES

- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- HOLLAND, J. L. Undergraduate origins of American scientists. *Science*, 1957, **126**, 433-437.
- HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245-254.
- KNAPP, R. H., & GOODRICH, H. B. *Origins of American scientists*. Chicago: Univer. Chicago Press, 1952.
- KNAPP, R. H., & GREENBAUM, J. J. *The younger American scholar: His collegiate origins*. Chicago: Univer. Chicago Press, 1953.
- THISTLETHWAITE, D. L. College environments and the development of talent. *Science*, 1959, **130**, 71-76. (a)
- THISTLETHWAITE, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183-191. (b)
- THISTLETHWAITE, D. L. College press and changes in study plans of talented students. *J. educ. Psychol.*, 1960, **51**, 222-234.
- UNITED STATES OFFICE OF EDUCATION. *Opening (fall) enrollment in higher education, 1958: Institutional data*. (Circular No. 544) Washington, D. C.: United States Government Printing Office, 1959.

(Received December 16, 1960)

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

August 1961

Number 4

## AN ANALYSIS OF THE FACTORIAL DIMENSIONS OF THE ACHIEVEMENT MOTIVATION CONSTRUCT

JAMES V. MITCHELL, JR.

*University of Texas*

The construct of achievement motivation has been the object of considerable research and discussion in recent years (Atkinson, 1958; McClelland, Atkinson, Clark, & Lowell, 1953), and it is only too apparent that there are wide differences of opinion about the adequacy of current efforts to measure that intriguing variable (Atkinson, 1960; Blake & Mouton, 1959). Much of that effort, of course, has been focused on the assessment of n Achievement by means of analyses of TAT protocols, a technique developed by McClelland et al. (1953) and now regarded as *sui generis* by its devotees. One adherent of this technique has even relegated all other methods of assessing achievement motivation to the limbo of "ad hoc" devices, and he admonishes all users of such devices not to assume that the theoretical conceptions and factual information resulting from projective analyses of n Achievement are equally pertinent to studies involving the "ad hoc" assessment devices (Atkinson, 1960, pp. 275-276). Others not quite so identified with the McClelland approach seem equally insistent that n Achievement scores from different instruments exhibit significant relationships to one another and to the common behavioral data they are supposed to measure. With discouraging regularity, however, investigators have reported the absence of relationship

between different measures of achievement motivation (Bendig, 1957; Himmelstein, Eschenbach, & Carp, 1958; Melikian, 1958), and attempts to predict actual behavioral criteria have often resulted in essentially negative findings, confusion, or vast conceptual reinterpretation (Birney, 1958; Lazarus, Backer, Broverman, & Mayer, 1957; Vogel, Baker, & Lazarus, 1958). It is readily apparent that there is a rather pronounced lack of consensus about both the conceptual and methodological aspects of the achievement motivation construct, and there is an urgent need for additional research that will help to provide structure for its currently amorphous state of existence.

The present investigation came about as a result of a larger study of the cultural and situational determinants of achievement motivation for a group of college students. For this larger study the measurement of achievement motivation was obviously a critical operation, and since the research literature provided ample evidence of the difficulties involved in this kind of assessment, it seemed imperative to select as many different measures of achievement motivation as would be necessary to assure a thorough representation of that domain and then to make a careful analysis of their correlates and factorial dimensions. It was early recognized that such

a procedure should result in findings that would have general theoretical significance and applicability, because they would help to shed light on whether achievement motivation was a unitary construct with invariable meaning, as many seem to regard it, or whether it referred in fact to several psychological dimensions, each with different implications and different empirical correlates. It is this task that defines the purpose of the present investigation.

### PROCEDURE

The following instruments were employed in the analysis to represent the domain of possible achievement related attitudes.

1. McClelland Test of Achievement Motivation. This instrument was administered and scored in accordance with the instructions given by McClelland in his book *The Achievement Motive* (McClelland et al., 1953). In scoring a representative sample of protocols provided in that book, the present investigator established that his scores correlated .92 with those reported by the test author.

2. A sentence completion test of achievement motivation. This test was developed by the investigator as a simplified projective measure of achievement motivation, with the thought that such a measure might well have more validity than an instrument with the scoring complexity of the McClelland test. The instrument consisted of 46 items, some developed by the author and some taken from experimental tests developed by other investigators. The items selected represented various degrees of achievement orientation, from strongly achievement oriented to complete lack of achievement orientation. The instrument was scored by applying to each item the very first criterion of the McClelland system for scoring the TAT protocols, viz., the presence or absence of an "achievement goal." According to McClelland, the existence of an achievement goal is indicated by any statement which implies "success in competition with some standard of excellence" (McClelland et al., 1953, p. 110). If such were expressed or implied in the completion of any item, that item was scored +1. If the item completion involved no achievement imagery of this type, it was scored 0. If the item completion expressed or implied an

evident lack of success or interest in competing with some standard of excellence, it was scored -1. The score on the test was simply the algebraic sum of the item scores.

3. An adjective checklist containing achievement related items. A few achievement related adjectives were added to those similarly oriented and already present in the Bills' Index of Adjustment and Values (Bills, Vance, & McLean, 1951), and for each of these adjectives the subject rated himself on a 5-point scale. The subject's achievement motivation score on this instrument was the total of the numerical ratings for the contributing adjectives. Corrected split-half reliability for this score was .84. Remaining items not contributing to this score served to mask the real purpose for which the instrument was being administered. Noncontributing items also served this same function for the two following instruments.

4. An inventory-type test of achievement motivation. This inventory contained items of the True-False variety, many of which were borrowed from the MMPI. Other items related to achievement motivation were added by the investigator. The corrected split-half reliability of this instrument was .75.

5. A multiple-choice questionnaire including achievement motivation items. The alternatives for each item in the questionnaire were serially ordered so as to define choices along a continuum representing greater or lesser degrees of the variable in question. The number of an alternative chosen thus defined a position on the continuum and the individual's score for that particular item. The corrected split-half reliability of this instrument was .83.

6. Taylor Manifest Anxiety scale (1953). This scale was included in the battery to clarify whether the imputed drive properties of manifest anxiety, as hypothesized by the disciples of Hull, would receive *de facto* verification from the factor analysis of these achievement motivation variables.

7. Level of aspiration measures. All subjects were members of the investigator's sections in elementary educational psychology. Before the first and last examinations of the semester all subjects were asked to record their levels of aspiration by answering the question: "What grade can you reasonably hope to attain on this examination?" These aspiration estimates were originally received in letter grade form and then which recognized + and - distinctions, e.g.,

C+, B-, etc. Actual grade received on the examination was also converted to the same 10-point numerical system. For each subject the numerical value of the actual grade received was subtracted from the numerical value of his level of aspiration estimate. This was done for both the first and last examinations, resulting in two indices which represented the extent to which the subject actually attained or approached his level of aspiration.

8. A measure of discrepancy between achievement and potentiality. The numerical grade averages for all subjects for the semester in which the study was conducted were computed, normalized, and converted into standard score form. The ACE scores for all subjects were similarly normalized, and converted into standard score form. A discrepancy score was computed for each subject by subtracting his standard score for the ACE from his standard score for his semester grade point average (GPA - ACE).

All the afore-mentioned data were secured from 131 female college students who were in a teacher training curriculum. The analysis was divided into two parts: (a) The scores from the various instruments were correlated with one another and with grade point average criteria, and (b) as a result of the insights gained from the first analysis, all of the individual items comprising the various instruments, some 29 variables in all, were then intercorrelated, subjected to a centroid factor analysis as described by Thurstone (1947), and subsequently rotated to orthogonal simple structure by means of Kaiser's varimax technique (1958).

#### RESULTS AND DISCUSSION

Table 1 shows the intercorrelations between measures of achievement motivation, ACE, and GPA. The GPA-ACE discrepancy score (last column) here functions as the criterion variable for achievement motivation. Immediately apparent are the differences in correlations between this criterion and the various measures of achievement motivation. The simple adjective checklist was the best predictor of the criterion. The achievement motivation inventory was significantly related to the criterion, but, unlike the

adjective checklist, had a significant negative correlation with ACE as well. The level of aspiration measures were related significantly to the criterion, but this may be partly artifactual in nature, resulting from the presence of a common component, grades, in the make-up of predictor and criterion. Manifest anxiety was quite unrelated to the criterion. Neither of the projective instruments (sentence completion and TAT n Achievement) was related to the criterion. Interestingly enough, however, the sentence completion test did have a significant and moderately high negative correlation with ACE and a lower but still significant negative correlation with GPA. These correlations, along with the negative correlation reported earlier for the achievement motivation inventory and ACE, re-emphasize the necessity for cautious empirical inquiry into the real dimensions and correlates of complex measures of complex constructs like achievement motivation.

The intercorrelations between the various measures of achievement motivation included in Table 1 show some interesting and often perplexing trends. The adjective checklist is significantly correlated with the achievement motivation inventory but not with the sentence completion test. The sentence completion test is, however, significantly correlated with the achievement motivation inventory. The pattern indicates some overlap between instruments that presumably measure the same variable, but hardly inspire confidence in the unidimensionality of the achievement motivation construct. The level of aspiration measures, besides being related to the criterion, are also related significantly to one or two other measures of achievement motivation. The TAT and manifest anxiety variables show little relationship to other measures.

TABLE 1  
INTERCORRELATIONS BETWEEN MEASURES OF ACHIEVEMENT MOTIVATION,  
ACE, AND GPA

Measures	Achievement Motivation Inventory	Sentence Completion	TAT n Achievement	Aspiration-Grade <sup>a</sup> (first test)	Aspiration-Grade <sup>a</sup> (last test)	Manifest Anxiety	ACE	GPA	GPA-ACE
Adjective Checklist	.40**	.13	-.07	.15	.02	-.15	-.02	.36**	.40**
Achievement Motivation Inventory		.35**	-.07	.07	-.20*	.08	-.23**	-.06	.18*
Sentence Completion			-.04	-.18*	-.06	.09	-.32**	-.17*	.09
TAT n Achievement Aspiration-Grade <sup>a</sup> (first test)				.13	.18* .26**	.01 .04	.11 .20*	.08 .48**	-.07 .27**
Aspiration-Grade <sup>a</sup> (last test)						.05	.19*	.48**	.27**
Manifest Anxiety							-.27**	-.11	.09
ACE								.44**	

<sup>a</sup> Direction of scoring reversed by subtracting each discrepancy from a constant; the higher scores represented the smaller discrepancies.

\* Significant at .05 level.

\*\* Significant at .01 level.

The whole pattern of interrelationships was one that suggested not only the multidimensionality of the putative achievement motivation construct, but also the probability that various measures of that construct might reflect quite different aspects of it and would therefore either be little correlated or selectively correlated with other measures and with the criterion. A factor analysis of all of the items of these various measures seemed admirably suited to shed light on such questions as these, and such an analysis was undertaken as the second step in the investigation.<sup>1</sup>

<sup>1</sup> The correlation matrix for the 29 individual achievement motivation items that were subjected to factor analysis has been deposited with the American Documentation Institute. Order Document No. 6739 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Table 2 shows the rotated factor matrix which resulted when the *individual items* of the achievement motivation measures were intercorrelated and factor analyzed. The items of the multiple-choice inventory, not included in the earlier analysis, were included in the present one to broaden the basis for analysis and comparison. Factorization was terminated when the factors isolated accounted for approximately 100% of the estimated common factor variance. A factor was not interpreted, however, unless the product of its two highest loadings exceeded two standard errors of a zero-order correlation for an *N* of the magnitude used in this study (Fruchter, 1954, p. 79). Two criteria were used for the rotational procedure: the varimax criterion (Kaiser, 1958), and the requirement that the final rotational solution achieve optimum factorial simplicity for the GPA-ACE criterion. After the first criterion was met, however, there was only one small ad-

TABLE 2

ROTATED FACTOR MATRIX FOR 29 INDICES OF ACHIEVEMENT MOTIVATION

Measures	I	II	III	IV	V	VI	VII
1. GPA-ACE	74	.01	.00	.05	-20	.08	.09
2. Aspiration-Grade Discrepancy <sup>a</sup> (first test)	30	-07	-32	.30	-05	.10	-14
3. Aspiration-Grade Discrepancy <sup>a</sup> (last test)	37	-04	-35	-11	-16	-06	-05
Adjective Checklist Items:							
4. Ambitious	25	.09	.33	.36	-22	.20	.26
5. Industrious	28	.13	.42	.43	-32	.17	.09
6. Hardworking	48	.13	.58	.26	-19	-09	.18
7. Studious	56	.07	.36	.40	.04	.04	.16
8. Competitive	11	.34	.08	.45	.11	-01	.08
9. Motivated	44	.15	.28	.60	-05	.03	.10
Achievement Motivation Inventory Items:							
10. I am more ambitious than the average person.	32	.04	.23	.29	-04	.31	.28
11. When I'm faced with difficult tasks, I generally "give up" too easily.	-06	.00	-05	-52	.05	.01	.14
12. I exert considerable effort to get good grades.	37	.12	.57	.03	.06	.28	.09
13. I wish I were more successful as a student.	-18	-33	.23	-20	.23	.06	-29
14. I try very hard to do well in anything I undertake.	29	.06	.18	.05	.12	.54	.12
15. I find that the standards I set for my own work are usually too high for me to attain.	.01	-31	.11	-15	.55	.17	-08
16. My parents exert considerable pressure on me to achieve good grades.	-27	.01	.07	.10	.49	.04	-08
17. I have strong motivation to succeed in my work.	49	.03	.40	.17	.20	.38	-17
18. I feel under constant pressure to do well in my school work.	.04	.02	.16	-01	.53	.04	.02
19. Manifest Anxiety	13	-36	-09	-11	.30	-17	.09
20. Sentence Completion	.09	-03	.33	-26	.14	.34	.07
Multiple-Choice Questionnaire Items:							
21. Do you feel that thus far you have really lived up to your parents' expectations for you?	33	.35	.07	.20	-41	.01	-17
22. Are your parents satisfied with your grades in college?	59	.27	.04	.27	-41	-11	-09
23. How hard do you work on your school-work in comparison with the typical student of your sex?	75	.05	.44	.00	.05	-02	-03
24. How important are <i>grades</i> to you?	34	-04	.48	-03	.06	.15	-02
25. Do you feel that your parents are satisfied with you as a person?	17	.65	.13	-02	-20	.01	-02
26. Are you satisfied with yourself as a person?	-01	.69	-04	.20	-01	.02	.04
27. Do you and your parents agree on "values"—what constitutes the really important things in life?	-03	.48	.24	-22	.11	.01	.24
28. How would you rate your study habits in comparison with the average student of your sex?	70	.15	.18	.13	-07	.10	.13
29. TAT n Achievement	-01	-02	-09	.05	.00	-06	-41

<sup>a</sup> Direction of scoring reversed by subtracting each discrepancy from a constant; the higher scores represented the smaller discrepancies.

ditional rotation that had to be made to meet the requirements of the second. The following factors were tentatively identified.

Factor I had its highest loadings for the GPA-ACE criterion and for two variables (23 and 28) which reflect the actual effort expended to do an effective academic job. Other variables have loadings to the extent that they also reflect an earnest application of effort to obtain tangible academic results. This factor might best be called Academic Motivation and Efficiency. It is interesting to note that the variables having the highest loadings on this criterion factor were objective self-rating indices, that the level of aspiration measures had moderate loadings on the factor, that the two projective measures (20 and 29) did not have loadings on this factor, and that neither manifest anxiety (19) nor self-satisfaction (26) had loadings on this important criterion factor.

Factor II had its two highest loadings for items relating to self-satisfaction and acceptance (25 and 26). It has negative loadings for manifest anxiety (19) and for the item expressing a desire to be more successful as a student (13). It seems appropriate to refer to this factor as Self-Satisfaction. It is interesting to note the lack of relationship between this Self-Satisfaction factor and the GPA-ACE criterion.

Factor III can best be interpreted in terms of its contrast with Factor I. It has no loading for the GPA-ACE criterion. It has negative loadings for the level of aspiration measures, indicative of its correlation with larger discrepancies between aspiration and achievement, as contrasted with the positive loadings of the first factor. It has a positive loading for the item expressing the desire for more success as a student, in contrast to the first factor's negative loading for the same

item (13). Its loading of .04 contrasts with the first factor loading of .59 for the item expressing parental satisfaction with the subject's grades (22). Its much smaller loadings for Items 23 and 28, earlier interpreted as reflecting actual willingness to expend effort to achieve results, must be contrasted with the corresponding very high loadings for the first factor. For most of the remaining variables there seems to be a similar tendency for low loadings to be associated with items referring explicitly to effective application to academic work, while the items with high loadings are more vague and generalized, less behavior oriented, and more susceptible to a positive response set. All the evidence, and particularly that already noted for Items 1, 2, 3, and 13, suggests that Factor III should be interpreted as Wish-Fulfillment Motivation. The factor seems to represent strong desire but little application or fulfillment.

It is interesting to compare the third factor loadings of the sentence completion test with its first factor loadings. The resulting contrast strongly suggests that projective test users must avoid jumping to conclusions about what level of experience or motivation is being tapped by a projective test. The appreciable loading of the sentence completion test on the Wish-Fulfillment factor, its lack of loading on the first factor, and the earlier reported correlation of -.32 between sentence completion and ACE, all suggest that this projective may tend to identify the less able student who yearns for high achievement but who, because of discouragement or inertia, cannot really buckle down to achieve results. This hypothesis also receives partial confirmation from the original correlations.

Factor IV has a high positive loading for the adjective "motivated" (9) and a high negative loading for the in-

ventory item indicating the subject's tendency to "give up" with difficult tasks (11). Several lesser loadings represent similarly achievement oriented variables. What most distinguishes this factor, however, is its lack of loadings for all items related strictly to academic motivation alone. Items having no loadings on this factor include two items concerned with willingness to work hard on schoolwork (12 and 23), one concerned with the importance of grades to the subject (24), and one concerned with the effectiveness of the subject's study habits (28). Because of this specific exclusion of academic motivation from what seems to be a generally positive achievement orientation, it seems logical to designate this factor as Nonacademic Achievement Motivation. Appropriately enough, this factor has no loading for the GPA-ACE criterion.

Factor V is notable for its high loadings for items concerned with feeling under parental pressure to do good academic work (16 and 18), aiming at standards that are too high to attain (15), but still not living up to parental expectations or achieving grades satisfactory to them (21 and 22). This factor seems most clearly to define an External Pressure to Achieve.

Factor VI, the most difficult factor to interpret, has a high loading for an item referring to the subject's willingness to try hard in all he undertakes (14) and lesser but still appreciable loadings for items reflecting a generalized desire to achieve (10, 12, and 17). Yet it is quite unrelated to the GPA-ACE criterion. Unlike the third factor, it has no loading for the level of aspiration measures (2 and 3) or for the item reflecting a desire for more success as a student (13). Tentatively we will call this factor Imputed Generalized Motivation without Attendant Effort. It is interesting to note that the sentence completion test again

loads on a factor that involves desire to achieve but little evidence of real or appropriately directed effort.

Factor VII should probably be interpreted as a residual or error factor of no importance. The product of its two highest loadings did not exceed the two standard error criterion. Its highest loading was for the TAT n Achievement test, whose appearance on this error factor is most probably due to low reliability. This same test had but one significant correlation in the entire battery, and this was low and perhaps attributable to chance.

An analysis of this type would seem to have several implications requiring the thoughtful consideration of researchers in this area. First, it is evident that the achievement motivation construct is *not* a unitary construct with invariable meaning, and that any attempt to regard it in this light is likely to lead to methodological confusion, conflicting results, and conceptual stagnation. In this study it was possible to isolate five or six relatively independent dimensions which represented different and psychologically significant aspects of the desire to achieve. It would therefore seem imperative for each research worker in this area to offer some empirical evidence of which of several possible varieties of achievement motivation he had chosen to employ in his investigation. To ignore the complexity of the achievement motivation construct is to render many results uninterpretable.

Secondly, it is apparent that tests and test items showing equivalent initial promise may turn out to differ widely in their ability to measure any given aspect of achievement motivation. Empirical surprises seem almost the rule rather than the exception. The TAT n Achievement test had the lowest communality of all 29 variables employed in the factor analysis, and

it loaded only on an error factor. Its one significant correlation in the original correlation matrix was a negative correlation with an item relating to ambition, and its correlation with the GPA-ACE criterion was -.07. On the other hand, two individual objective items (23 and 28) had very high loadings on the first factor and were correlated .53 and .52, respectively, with the GPA-ACE criterion. This is a notable result, not only because such correlations are even difficult to achieve with whole tests, but also because they were achieved with objective, multiple-choice, self-rating items that focus on concrete, easily observed behaviors that are a part of everyday experience.

Thirdly, tests and test items may differ widely in terms of which particular aspects of the desire to achieve complex they actually measure. For example, the sentence completion test had a .09 loading on the first factor, a correlation of .09 with the GPA-ACE criterion, and significant correlations of -.32 and -.17 for the ACE and GPA, respectively. There is certainly no evidence of its validity as a measure of achievement motivation here, and if GPA had been used as a criterion, results would be directly contrary to hypothesis. But the complete analysis revealed that this test was loaded on the Wish-Fulfillment factor, and its loading there gave credence to Lazarus' (Lazarus et al., 1957) contention that there is no one-to-one relationship between the expression of a need in fantasy and its actualization in behavior. Indeed, the fantasy may even serve as a substitute for overt striving. Analyses of fantasy, then, may not always be "depth" analyses; they may actually reveal relatively superficial aspects of behavior and personality.

Certain objective items also turned out to measure quite different aspects

of the desire to achieve complex than had been originally anticipated. It had been anticipated, for instance, that Items 8, 11, and 18 would be highly related to the criterion. As it developed, they were not at all related to the criterion nor to the first factor; they were, however, loaded on other factors. Items 22 and 28, on the other hand, were highly related to the GPA-ACE criterion, having correlations of .50 and .52, respectively, but their patterns of loadings in the factor matrix were quite different. Here, too, cautious analysis is needed to ascertain just what aspect of the desire to achieve complex is actually being measured.

The importance of such analysis is particularly apparent in comparing the lack of correlation between some of the instruments represented in Table 1 with the relatively high loadings of the factor matrix and the relatively high interitem correlations of the original correlation matrix. A comparison of this kind furnishes compelling evidence that intertest correlations can be attenuated by items representing different factorial dimensions or by same-test items actually counteractive in their influence.

Fourthly, it is not only important to acknowledge that different tests and test items may represent different aspects of the desire to achieve complex; it is equally important to recognize that the same test or test item may represent several different aspects of the desire to achieve complex. The adjective "industrious" (9), for instance, has loadings on four different factors and could represent four distinctly different attitudes toward achievement. There are several other items (e.g., 9, 17, and 23) that are loaded on two or three factors. It is apparent that identical responses can have very different meanings for different subjects.

Finally, there were certain by-products of this investigation that are worthy of note despite their peripheral status. The pattern of loadings for manifest anxiety did not lend confirmation to the hypothesized drive attributes of that variable, but it must be acknowledged that this investigation was not designed as a critical test of that issue. It is also interesting to note that the item relating to self-satisfaction (25) was correlated .00 with the GPA-ACE criterion and .09 with GPA alone. The Self-Satisfaction factor had similarly insignificant and negligible loadings for items concerned with strictly scholastic motivation (7, 17, 23, and 24). Apparently the self-concepts of our college students are not nearly as influenced by matters relating to academic motivation and achievement as we are sometimes inclined to believe.

### SUMMARY

The purpose of this study was to determine whether "achievement motivation" was a unitary construct with invariable meaning or a complex of relatively independent dimensions with different implications and different empirical correlates. Twenty-nine indices of achievement motivation were subjected to factor analysis. Six factors were identified: Academic Motivation and Efficiency (the only factor highly predictive of actual academic performance), Wish-Fulfillment Motivation, Nonacademic Achievement Motivation, Self-Satisfaction, External Pressure to Achieve, and Imputed Generalized Motivation without Attendant Effort. In addition to demonstrating that there are several varieties of achievement motivation, the results also indicated that: tests and test items differed widely in their ability to measure any of the several facets of this complex variable, identical responses could have very different

meanings for different subjects, self-rating procedures excelled projective methods in predicting actual academic performance, and the projectives were either loaded on wish-fulfillment or error factors.

### REFERENCES

- ATKINSON, J. W. *Motives in fantasy, action, and society: A method of assessment and study*. Princeton: Van Nostrand, 1958.
- ATKINSON, J. W. Personality dynamics. *Annu. Rev. Psychol.*, 1960, **11**, 255-290.
- BENDIG, A. W. Manifest anxiety and projective and objective measures of need achievement. *J. consult. Psychol.*, 1957, **21**, 354.
- BILLS, R. E., VANCE, E. L., & McLEAN, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257-261.
- BIRNEY, R. C. The achievement motive and task performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 133-135.
- BLAKE, R. R., & MOUTON, J. S. Personality. *Annu. Rev. Psychol.*, 1959, **10**, 203-232.
- FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
- HIMELSTEIN, P., ESCHENBACH, A. E., & CARPENTER, A. Interrelationships among three measures of need achievement. *J. consult. Psychol.*, 1958, **22**, 451-452.
- KAISSER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
- LAZARUS, R. S., BACKER, R. W., BROVERMAN, D., & MAYER, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559-577.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MELIKIAN, L. H. The relationship between Edwards' and McClelland's measures of achievement motivation. *J. consult. Psychol.*, 1958, **22**, 296-298.
- TAYLOR, A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
- THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.
- VOEGL, W., BAKER, R. W., & LAZARUS, R. S. The roles of motivation in psychological states. *J. abnorm. soc. Psychol.*, 1958, **56**, 105-112.

# CATEGORIZATION AND RELATED VERBALIZATIONS IN DEAF AND HEARING ADOLESCENTS<sup>1</sup>

SOLIS L. KATES, WILLIAM W. KATES, JAMES MICHAEL,  
AND TERRENCE M. WALSH

*Clarke School for the Deaf, Northampton, Massachusetts,  
and University of Massachusetts*

Studies dealing with the perceptual and conceptual capacities of deaf children have suggested that children deaf from birth fail to develop the ability to deal with abstract material in a fashion comparable to hearing children (Larr, 1955; Myklebust, 1953; Oleron, 1950; Templin, 1950). Other studies, however, have shown that the differences between the thinking processes of deaf and hearing children do not correspond to the differences between the concreteness of the aphasics and the abstract functioning of normal subjects (Oleron, 1953) but correspond much more closely to the differences between a younger and older group of hearing children (Clarke School, 1940). When the level of verbalization required is within the range of vocabulary and experience of the deaf subject, deaf subjects perform as well as hearing subjects on abstract as well as less abstract tasks (Rosenstein, 1959).

There are three major purposes of this experiment. First, we are interested in investigating a specific aspect of the cognitive processes of deaf and hearing children, that of categorization. Categorization, "to render discriminably different things equivalent," forms one of the bases for all other cognitive activities (Bruner, Goodnow, & Austin, 1956, p. 1). Our second major concern is in distinguishing between the process of categorization on the one hand, and the process

of verbalization on the other, and to judge the adequacy of categorization independently of the adequacy of verbalization. The process of learning may involve both learning a new verbal attribute and then learning the non-verbal category that goes with it as well as learning a verbal attribute to go with a nonverbal category already formed (Brown, 1958). Our third concern is in attempting to separate the effects of deafness from the cognitive effects of age and achievement level.

The general hypothesis of this study was: There will be no differences in categorization or verbalization found for one comparison of deaf subjects with hearing subjects of matched age and IQ that will also be found for the comparison of these deaf subjects with hearing subjects matched on school achievement and IQ.

## METHOD

*Subjects.* One group consisted of five girls and three boys at the Clarke School for the Deaf in Northampton, Massachusetts, representing the total number of graduating seniors with very profound hearing loss. Their training was by the oral method of instruction. Their average IQ on the WAIS (both verbal and performance subscales) was 102.63; *SD*, 6.93. On the Stanford Achievement Test, their average achievement was equivalent to the 7.79 grade level. Their mean age was 18.09 years; *SD*, 0.33.

Our second group was composed of eight junior high school students, five girls and three boys. Their average Stanford Achievement level was equivalent to the 7.68 grade level. Their average IQ, based on WISC and Terman-McNemar, was 104.63; *SD*, 5.05. Their average age was 14.2 years.

Our third group included eight high school seniors, five girls and three boys.

<sup>1</sup>This study supported in part by Grant No. 549 from the Office of Vocational Rehabilitation, United States Department of Health, Education, and Welfare.

Their average age was 17.97 years;  $SD$ , 0.50. Their average IQ, based on WAIS and Terman-McNemar, was 106.38;  $SD$ , 5.50. These subjects were achieving at an average level for their high school class, and were expected to graduate without difficulty. The high school and junior high school students were screened carefully in consultation with their teachers and counselors, and no subject was included with any atypical personality and physical characteristics.

There were no significant differences between the deaf and the junior high school subjects on achievement level and IQ, and, similarly there were no significant differences between the deaf and the high school students on age and IQ. The size of the sample makes it necessary for this study to be regarded as exploratory and, of course, subject to replication.

*Categorization and Verbalization Tests.* The Goldstein-Scheerer (1941) Object Sorting Test was individually administered. This test consists of 33 everyday objects including such things as a real cigar, a real pair of pliers, a toy hammer, etc. (Rapaport, Gill, & Schafer, 1945).

*Procedure.* The Object Sorting Test was administered in three parts: the free, active, and compliant sections. In the free sorting section, the subject was asked to put all the objects into any number of groups and then tell why the objects in each group belonged together. In the active section, the subject was asked to put all the objects together that belonged with a preselected sample object and give his reasons for the categorization. There were eight "target" objects, the first selected by the subject and rest presented by the experimenter. The compliant section involved the presentation of 12 different preselected groupings of objects and asking each subject to give the reasons for the groupings. All three sections of the test were administered in one session and all subjects received the free, active, and compliant sections in order. Written instructions were used for the deaf subjects and the same instructions were given orally to the hearing subjects. There was no doubt on the part of the experimenters that all verbalizations of all subjects were well understood.

*Scoring.* A scoring system for categorizations was devised, based partially on that used by Rapaport et al. (1945, pp. 401-406). Classification of verbalizations were based upon those used by Rapaport as well as upon the criteria for types of rules of verbalization set up by Bruner et al. (1956, pp.

4-7). Two independent scorers agreed on 95% of a sample of 350 items scored independently. Copies of the complete scoring system are available upon request.

*Statistical Procedure.* Our statistical analysis compared separately the deaf subjects with each one of the two control groups; the question of the overall differences among the three groups was not a concern of this study.

To make some data comparable from subject to subject, proportions were calculated for measures such as adequacy of categorization and verbalization. Estimate tests of homogeneity of variance were performed and heterogeneous data were converted to  $T$  scores (Edwards, 1954). All variances were then found to be homogeneous.

## RESULTS AND DISCUSSION

*Prediction 1.* The deaf subjects will have a similar number of divisions and categories as the hearing subjects matched on achievement and IQ but significantly more categories and divisions than the hearing subjects matched on age and IQ. Furthermore, the deaf subjects will not differ in the number of failures from either hearing group. Divisions are defined as the number of groupings and/or single objects that the 33 objects are divided into in the free sorting section. The number of categories is equal to the number of groupings with two or more nonidentical objects included. A failure consists of an object that is not grouped with another object or of an absolute identity class such as two sugar cubes.

*Results.* The results, in general, support this prediction. Deaf subjects have the same number of divisions, categories, and failures as both groups of hearing subjects (Table 1), and thus appear as effective as hearing subjects in the two control groups in reducing the complexity of this experimental world.

*Prediction 2.* The deaf subjects will categorize as adequately but will not

TABLE 1

*t* TEST RESULTS FOR COMPARISONS BETWEEN DEAF AND CONTROL GROUPS MATCHED ON AGE AND IQ ( $C_1$ ) AND ON STANFORD ACHIEVEMENT AND IQ ( $C_2$ )

Comparison	Group	M	SD	<i>t</i> <sup>a</sup>
Divisions	$C_1$	8.75	1.79	0.298
	Deaf	9.00	1.32	
	$C_2$	7.88	1.60	1.460
Categories	$C_1$	7.13	1.05	0.182
	Deaf	7.25	1.39	
	$C_2$	6.50	1.22	1.070
Failures	$C_1$	2.13	1.97	0.00
	Deaf	2.13	0.93	
	$C_2$	2.25	1.92	0.155
Concrete verbalization <sup>b</sup>	$C_1$	0.080	0.078	1.418
	Deaf	0.158	0.123	
	$C_2$	0.212	0.115	0.857
Functional verbalization <sup>b</sup>	$C_1$	0.214	0.222	0.404
	Deaf	0.176	0.113	
	$C_2$	0.275	0.123	1.571
Category width with failures	$C_1$	5.00	1.16	2.239*
	Deaf	3.65	1.10	
	$C_2$	4.13	0.72	0.968
Category width without failures	$C_1$	5.20	1.03	2.482*
	Deaf	3.79	1.09	
	$C_2$	4.51	0.93	1.324
Formal verbalizations used inadequately <sup>c</sup>	$C_1$	52.33	11.07	0.031
	Deaf	52.18	6.15	
	$C_2$	47.87	5.88	1.340
Inclusive verbalization <sup>d</sup>	$C_1$	47.35	9.18	0.898
	Deaf	52.33	11.45	
	$C_2$	50.37	7.42	0.380

<sup>a</sup> The *t* value is the result of the comparison made of the deaf group mean with the control group mean just opposite the corresponding *t* value.

<sup>b</sup> Raw scores converted to proportion of total number of verbalizations.

<sup>c</sup> Proportion of formal inadequate to total inadequate converted to *T* scores.

<sup>d</sup> Proportion of inclusive verbalization to total number of verbalizations converted to *T* scores.

\* Significant at  $p < .05$  level.

have as many adequate verbalizations as the hearing subjects matched on age and IQ. Furthermore, the deaf subjects will have the same number of

adequate verbalizations as the group matched on achievement and IQ. An adequate categorization is one in which all the objects included are relevant to each other independent of the verbalization, and there are no relevant objects present in the sample that are not included in the grouping. An adequate verbalization is defined as one that covers correctly the group sorted. An inadequate verbalization may be false, or may not take into account all the objects included in the categorization, or may take into account all objects included in the categorization but pertain to objects outside the categorization for which the verbalization is given.

*Results.* The results support this prediction. Analyses of variance performed on the proportion of adequate categorizations to the opportunities for categorization show that there are no differences between the deaf subjects and either of the two hearing groups (Tables 2 and 3). When the task of grouping equivalent objects together is judged independently of the verbalization of the reason for that grouping, deaf subjects perform as well as hearing subjects of their own age and IQ. Furthermore, the deaf subjects have significantly fewer adequate verbalizations than the hearing group matched on age and IQ ( $F = 6.58$ ,  $p < .05$ ) but there are no differences between the deaf group and the hearing group matched on achievement and IQ (Tables 2 and 3). Thus, in terms of adequacy of verbalization, the deaf subjects seem to resemble the younger hearing subjects.

*Prediction 3.* The deaf subjects will have a greater number of adequate categorizations accompanied by inadequate verbalizations than the control group matched on age and IQ while not differing from the hearing group matched on achievement and IQ.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE  
DEAF GROUP AND THE TWO CONTROL  
GROUPS

Variable	Group	<i>M</i>	<i>SD</i>
Adequacy of categorization	Deaf	0.713 <sup>a</sup>	0.200
	C <sub>1</sub> <sup>b</sup>	0.755	0.187
	C <sub>2</sub> <sup>c</sup>	0.674	0.149
Adequacy of verbalization	Deaf	0.624 <sup>d</sup>	0.224
	C <sub>1</sub>	0.774	0.185
	C <sub>2</sub>	0.582	0.197
Adequate categorization— inadequate verbalization	Deaf	52.48 <sup>e</sup>	8.86
	C <sub>1</sub>	45.32	6.54
	C <sub>2</sub>	52.83	8.25
Formal generic verbalization	Deaf	52.70 <sup>e</sup>	8.92
	C <sub>1</sub>	51.99	9.54
	C <sub>2</sub>	45.37	8.95
Total formal verbalization	Deaf	51.55 <sup>e</sup>	8.90
	C <sub>1</sub>	52.53	10.40
	C <sub>2</sub>	45.82	8.08
Proneness to change	Deaf	48.63 <sup>e</sup>	8.56
	C <sub>1</sub>	52.05	9.80
	C <sub>2</sub>	50.65	7.74

<sup>a</sup> Proportion of the number of adequate categories to total opportunities for categorization.

<sup>b</sup> C<sub>1</sub> is the control group matched with the deaf on age and IQ.

<sup>c</sup> C<sub>2</sub> is the control group matched with the deaf on Stanford Achievement and IQ.

<sup>d</sup> Proportion of the number of adequate verbalizations to the number of categorizations.

<sup>e</sup> Converted to *T* scores.

*Results.* The analyses were conducted on standard scores converted to *T* scores (Tables 2 and 3). The results indicate that the deaf subjects have a significantly greater number of adequate categorizations accompanied by inadequate verbalizations than the control subjects matched on age and IQ ( $F = 5.47$ ;  $p < .05$ ). But there were no differences between the deaf subjects and those matched with the deaf on achievement and IQ. The results support our prediction. It may be pointed out that the younger control group had significantly more ade-

TABLE 3

ANALYSIS OF VARIANCE RESULTS FOR COMPARISONS BETWEEN DEAF AND THE CONTROL GROUP MATCHED ON AGE AND IQ (C<sub>1</sub>) AND ON STANFORD ACHIEVEMENT AND IQ (C<sub>2</sub>)

Variable	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Adequacy of categorization <sup>a</sup>				
C <sub>1</sub> Between treatment groups	1	0.014	0.014	0.280
Error	14	0.699	0.050	
C <sub>2</sub> Between treatment groups	1	0.013	0.013	0.317
Error	14	0.573	0.041	
Adequacy of verbalization <sup>b</sup>				
C <sub>1</sub> Between	1	0.270	0.270	6.585*
Error	14	0.573	0.041	
C <sub>2</sub> Between	1	0.021	0.021	0.438
Error	14	0.677	0.048	
Adequate categorization—inadequate verbalization <sup>c</sup>				
C <sub>1</sub> Between	1	410.06	410.06	5.474*
Error	14	1048.76	74.91	
C <sub>2</sub> Between	1	0.96	0.96	0.009
Error	14	1544.34	110.31	
Formal generic verbalization <sup>d</sup>				
C <sub>1</sub> Between	1	5.98	5.98	0.003
Error	14	3064.02	218.86	
C <sub>2</sub> Between	1	645.55	645.55	3.417
Error	14	2644.15	188.82	
Total formal verbalization <sup>e</sup>				
C <sub>1</sub> Between	1	11.59	11.59	0.619
Error	14	2622.46	187.32	
C <sub>2</sub> Between	1	385.90	385.90	3.078
Error	14	1755.12	125.37	
Proneness to Change <sup>f</sup>				
C <sub>1</sub> Between	1	93.61	93.61	0.785
Error	14	1670.42	119.32	
C <sub>2</sub> Between	1	32.46	32.46	0.382
Error	14	1190.63	85.05	

<sup>a</sup> Raw scores converted to proportion of adequate categorizations to total number of categorizations.

<sup>b</sup> Raw scores converted to proportion of adequate verbalizations to total number of verbalizations (number of divisions plus 8 active plus 12 compliant).

<sup>c</sup> Raw scores converted to *T* scores.

<sup>d</sup> Proportions of formal generic verbalizations to total number of verbalizations converted to *T* scores.

<sup>e</sup> Proportions of total formal verbalizations to total number of verbalizations converted to *T* scores.

<sup>f</sup> Significant at  $p < .05$  level.

quate sortings accompanied by inadequate verbalizations than the older control group. Our results suggest that deafness holds back the process of correctly relating verbal attributes to nonverbal categories but that this retardation is similar to that shown by younger hearing subjects when they are compared to older hearing subjects.

*Prediction 4.* The deaf subjects will have more functional, more total concrete, fewer formal, fewer formal generic verbalizations than the control group matched on age and IQ but the same number as the control group matched on achievement and IQ. A functional verbalization specifies a rule of categorization based upon the sharing of a common function. Total concrete verbalizations include the syncretistic, fabulatory, chain, symbolic, split narrow verbalization as defined by Rapaport et al. (1945), and formal specific verbalization type. Formal verbalizations are defined as verbalizations specifying a rule of grouping based on the sharing of an intrinsic property ("all red," or "all kitchen utensils") among the objects in the group (Bruner et al., 1956). A formal generic verbalization specifies a rule of grouping based upon the sharing of highly complex and class-like attributes ("all metal").

*Results.* There are no differences between the deaf and either of the hearing control groups with respect to any of the verbalization types discussed above (Tables 1, 2, and 3). This prediction is not supported. However, the group matched with the deaf on achievement and IQ had significantly more "total concrete" verbalizations than the older control group. This result indicates that the test does differentiate, at least in regard to concrete verbalizations, levels of verbalization due to age and achievement differences in hearing subjects. On the whole these

concrete verbalizations represent an inability to utilize a conjunctive rule of grouping to explain a categorization. Further, they can also be described as "private" verbalizations (McGaughran & Moran, 1956). Deaf subjects on these tasks are no more concrete verbally, do not demonstrate any greater inability to utilize conjunctive rules of grouping, and utilize no more private verbalizations than the two control groups.

There is little evidence and theory concerning the following issues, and they are stated in the form of a question.

Are there any differences between the deaf subjects and either control group in proneness to change, in category width, in number of formal verbalizations used inadequately, and in the number of too inclusive verbalizations? Proneness to change is measured by the exclusion of objects in the active sorting which were included with these target objects in free sorting or by the inclusion, with the target objects in the active sorting, of objects not included with them in the free sorting. Category width is measured by the number of objects included with the target objects in the active section, both counting and not counting failures.

The data were converted to *T* scores and analysis showed that there are no differences between deaf subjects and either of the two control groups in overall proneness to spontaneous change (Tables 2 and 3). These results are in contrast to other results which have suggested that deaf subjects are more rigid than hearing children (McAndrew, 1948). The question of the rigidity of the deaf is not a settled one, since our evidence points to no difference between groups when spontaneous behavior is examined.

Our results show that the deaf sub-

jects have narrower categories than the control group matched on age and IQ but do not differ in category width from the hearing subjects matched on achievement and IQ (Table 1). The use of the narrower categories by the deaf subjects who, at the same time have the same number of categories as the hearing groups, may indicate a greater difficulty in organizing the world and in generalizing rules of grouping to a wide array of objects. This, of course, would also be true of the younger group of hearing subjects from whom the deaf did not differ in category width. However, with respect to this small array of familiar objects, the use of narrower categories might be more appropriate and more predictive than the use of wider categories. The problem of the relation of deafness to category width seems to be one that deserves future attention.

The two proportions of formal inadequate verbalizations and of inclusive verbalizations to the total number of verbalizations, respectively, were converted to *T* scores and were compared for the three groups. The deaf did not differ significantly from either control group in formal inadequate verbalizations and in inclusive verbalizations, respectively (Table 1).

#### SUMMARY

This study has investigated the categorization and verbalization processes of 24 adolescents, 8 of whom were profoundly deaf. The remaining subjects comprised two hearing groups, one group matched with the deaf subjects on age and IQ and the other group matched on Stanford Achievement and IQ. The Goldstein-Scheerer Object Sorting Test was administered and the deaf subjects were separately compared with each of the two control groups.

Our conclusions are as follows:

1. Deaf subjects categorize, on this task, as adequately as the hearing subjects. They have the same number of divisions, failures, and categories as the hearing subjects. On categorization tasks judged independently of verbalization, deaf subjects perform as well as hearing adolescents.

2. Deaf subjects have more inadequate verbalizations than hearing subjects and more adequate categorizations accompanied by inadequate verbalizations than hearing subjects matched on age and IQ. These differences are not true for the comparison of the deaf subjects and the hearing subjects matched on achievement and IQ. It is proposed that deafness retards learning of the correct relation between the proper verbal attribute and the nonverbal category but that this retardation lies along a normal developmental progression.

3. Deaf subjects do not differ from hearing subjects in the type of verbalization or in the developmental level of the verbalizations used, or in their spontaneous changes of categorization throughout the test.

4. Deaf subjects have narrower categories than hearing adolescents of their own age and IQ though they do not differ from hearing subjects of the same achievement and IQ.

5. There were no differences found between deaf and hearing subjects that indicated a definite shift in categorization ability distinct from the effects of age and achievement.

#### REFERENCES

- BROWN, R. *Words and things*. Glencoe, Ill.: Free Press, 1958.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- CLARKE SCHOOL FOR THE DEAF, Clarence W. Barron Research Department, Psychological Division. Studies in the psychol-

- ogy of the deaf. *Psychol. Monogr.*, 1940, **52**(1, Whole No. 232).
- EDWARDS, A. L., *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- GOLDSTEIN, K., & SCHEERER, M., Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).
- LARR, A. L. An experimental investigation of the perceptual and conceptual abilities of children in residential schools for the deaf. Unpublished doctoral dissertation, Syracuse University, 1955.
- McANDREW, H. Rigidity and isolation: A study of the deaf and the blind. *J. abnorm. soc. Psychol.*, 1948, **43**, 476-494.
- MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups.
- J. abnorm. soc. Psychol.*, 1956, **52**, 43-51.
- MYKLEBUST, R., Towards a new understanding of the deaf child. *Amer. Ann. Deaf*, 1953, **98**, 345-357.
- OLERON, P. A study of the intelligence of the deaf. *Amer. Ann. Deaf*, 1950, **95**, 179-195.
- OLERON, P. Conceptual thinking of the deaf. *Amer. Ann. Deaf*, 1953, **46**, 299-315.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
- ROSENSTEIN, J. Perception and cognition in deaf children. Unpublished doctoral dissertation, Washington University, 1959.
- TEMPLIN, M. The development of reasoning in children with normal and defective hearing. Minneapolis: *Univer. Minnesota Press*, 1950.

(Received September 6, 1960)

# CROSS-VALIDATION GROUPS, EXTREME GROUPS, AND THE PREDICTION OF ACADEMIC ACHIEVEMENT

LEONARD J. KAZMIER<sup>1</sup>

*University of Notre Dame*

In two parallel studies, scores on the 15 personality variables of the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954) have been compared for over- and underachieving college students (Gebhart & Hoyt, 1958; Krug, 1959). In both studies, extreme over- and underachievers were selected from a larger sample of male college students, so that the middle range of achievers was eliminated from the study. Accordingly, analyses of variance were performed to test for the significance of differences between the extreme groups, but validity coefficients were not computed. Though the scales that were found to discriminate between subjects who over- or underachieved in relation to measured aptitude were not identical, the results of the two studies did agree to a considerable extent. In the Gebhart and Hoyt study, overachievers scored higher than underachievers on the Achievement, Order, Intracception, and Consistency scales, and significantly lower on the Nurturance, Affiliation, and Change scales. In the study by Krug, overachievers scored significantly higher on the Achievement, Order, and Endurance scales and significantly lower on Affiliation and Heterosexuality. The .05 level was defined as significant in both studies.

There has come to be considerable concern about the widespread use of extreme groups in hypothesis testing. As McNemar has indicated (1960), such tests of significance are misleading and may represent only a small degree of relationship between the two

variables. The primary purpose of the present study was to investigate the relationship between the EPPS scales and a criterion of academic achievement when scores of all subjects are utilized, not just those of selected extreme groups. The correlation technique was used throughout because it provides a measure of the degree of relationship as well as its significance. To ascertain that the measured relationships of the several scales with the criterion are not due to chance factors, a cross-validation group was included in the study.

## PROCEDURE

The sample consisted of 140 students enrolled in two sections of introductory psychology at Ohio State University during the Spring Quarter of 1960. Both sections were taught by the same instructor, the author of the present report. For use as a predictor, six paragraphs of material, including a total of 36 questions on the contents thereof, were chosen from old forms of the Ohio State Psychological Examination (Buros, 1959, pp. 494-495). The criterion for success in the psychology course was the total number of points accumulated at the end of the quarter from three departmental examinations and four class quizzes.

In order to achieve comparability of the experimental and cross-validation groups, all class cards for the subjects were segregated by sex, and then categorized according to college year. Within the college year the cards were arranged in order of percentile scores on the Ohio State Psychological Examination. Seven students for whom these scores were not available were arbitrarily placed at the end of their respective college-year groups, after the highest percentile scores. Then the overall sample was subdivided into two groups of 70 subjects each by alternatively putting one card in the experimental group and the following card in the cross-validation group. Comparability of the two groups in terms of

<sup>1</sup> Research conducted while the author was at Ohio State University.

their standing on the reading comprehension test, the criterion, and the 15 EPPS scales was ascertained.

The correlation of each of the 15 scales with the criterion of success was determined for each of the groups and for the entire sample, and the Wherry-Doolittle test selection technique (Garrett, 1958, pp. 426-440) was then applied to the experimental group of subjects in order to determine the best selection and weighting of EPPS scales for maximum predictiveness. The applicability of this result to the cross-validation group was investigated by computing a composite  $r$  for that group by using the scales and weights derived from the experimental group.

### RESULTS

The split-half reliability coefficient (odd vs. even) for the reading comprehension test is .813, while the coefficient of correlation between this predictor and the criterion of total points in the psychology course is .593, which is significant at the .001 level for the 140 subjects involved in this study.

TABLE 1  
COMPARISONS BETWEEN THE EXPERIMENTAL  
AND CROSS-VALIDATION GROUPS

Measure	Experimental Group (N = 70)		Cross-Validation Group (N = 70)		
	M	SD	M	SD	F
Points in psychology course	259.00	39.25	260.10	36.05	— <sup>a</sup>
Predictor Score	17.18	6.22	16.84	5.98	—
EPPS Scales:					
Achievement	13.67	3.94	14.23	4.75	—
Deference	11.41	3.71	12.21	3.67	1.64
Order	9.44	4.17	10.67	3.90	3.24
Exhibition	14.34	3.33	14.34	3.65	—
Autonomy	12.86	4.20	12.66	5.04	—
Affiliation	16.26	4.38	16.57	4.49	—
Intraception	17.26	4.88	17.34	4.78	—
Succorance	11.38	5.25	11.00	4.13	—
Dominance	15.58	4.68	14.41	5.23	1.95
Abasement	14.31	4.99	14.70	5.67	—
Nurturance	15.66	5.18	16.18	4.81	—
Change	17.57	4.78	16.90	4.55	—
Endurance	13.63	5.67	13.08	5.27	—
Heterosexuality	14.97	5.08	15.08	5.58	—
Aggressiveness	11.00	4.62	10.44	4.54	—

<sup>a</sup> F < 1.00.

TABLE 2  
CORRELATION OF THE EPPS SCALES WITH  
THE CRITERION OF SUCCESS  
IN PSYCHOLOGY

Scale	Experimental Group (N = 70)	Cross Validation Group (N = 70)	Total Sample (N = 140)
Achievement	.113	.191	.153*
Deference	.013	-.109	-.043
Order	.029	-.214*	-.080
Exhibition	.027	.157	.092
Autonomy	.073	-.040	.013
Affiliation	-.139	-.161	-.149*
Intraception	.022	.099	.059
Succorance	-.205*	.078	-.087
Dominance	.209*	.226*	.213**
Abasement	-.054	-.068	-.060
Nurturance	-.167	-.163	-.164*
Change	-.244**	.035	-.114
Endurance	.088	-.061	.018
Heterosexuality	.192	.037	.114
Aggressiveness	.115	-.034	.043

\* p < .10.

\*\* p < .05.

That the procedure used to divide the subjects into experimental and cross-validation groups did result in comparability of the groups in many important respects is indicated by the data in Table 1, which presents the mean scores and standard deviations for the two groups on the reading comprehension test, the criterion, and the 15 EPPS scales. Not one of the obtained  $F$  ratios is significant at the .05 level.

The correlation coefficients between the criterion of success in the psychology course and each of the 15 scales of the EPPS are indicated in Table 2 for the experimental and cross-validation groups as well as for the total sample. For the experimental group, the correlation coefficients for Succorance and Dominance were significant at the .10 level, while that for Change was significant at the .05 level. The cross-validation group yielded a somewhat different pattern of correlations, re-

sulting in significance at the .10 level for the Order and Dominance scales. For the total sample of subjects, the correlation coefficients for Achievement, Affiliation, and Nurturance were significant at the .10 level while that for Dominance was significant at the .05 level.

It would appear that the desired stability of relationship of the EPPS scales with the criterion is lacking in this study. For example, of the three scales whose relationship with the criterion was significant at the .10 level or better in the experimental group, only one scale "held up" for the cross-validation group, while the algebraic signs of the correlation coefficients for the other two groups actually were reversed. In order to verify this conclusion statistically, the Wherry-Doolittle technique was utilized to ascertain the best selection of scales for the experimental group and the appropriate beta weights to be used in conjunction with each scale. Stability of the scales is then indicated by the extent to which these beta weights apply for use with the cross-validation group. As indicated by Table 3, for the experimental group it was found that the use of the Change, Succorance, and Heterosexuality scales results in the maximal coefficient of multiple correlation corrected for shrinkage. Addition of the fourth best scale, Abasement, results in a decrease in the magnitude of the correlation coefficient.

The beta weights obtained by the Wherry-Doolittle test selection method are -.252 for Change, -.233 for Succorance, and .181 for Heterosexuality. Utilizing these weights derived for the experimental group and the correlation coefficients of the chosen scales with the criterion in the cross-validation group results in a composite coefficient of correlation for the cross-validation group of -.058.

TABLE 3

BEST SELECTION OF EPPS SCALES FOR THE EXPERIMENTAL GROUP BY THE WHERRY-DOOLITTLE METHOD

Scales	$s\bar{R}$
Change	.244
Change + Succorance	.313
Change + Succorance + Heterosexuality	.345
Change + Succorance + Heterosexuality + Abasement	.331

Thus, the validity evidenced for the EPPS scales in the experimental group did not at all hold up in the cross-validation group, in that the scales and weights that resulted in a multiple  $R$  of .345 for the experimental group yielded a slightly negative composite coefficient for the second group.

#### SUMMARY

Previous studies have reported predictiveness for certain scales of the Edwards Personal Preference Schedule (EPPS) in regard to academic over- and underachievement when the analysis of variance technique was applied to selected extreme groups. The purpose of the present study was to test out the EPPS in a situation in which no subjects are eliminated from the study, the correlation technique is used, and both experimental and cross-validation groups are included.

The selection of the best EPPS scales through the use of the Wherry-Doolittle test selection technique resulted in a coefficient of multiple correlation of .345 for the experimental group. However, when these scales were used for the cross-validation group, the value of the correlation coefficient degenerated to -.058.

The results of the study do not necessarily indicate a lack of usefulness for the EPPS, but they do highlight

the shortcomings of many of the research studies which have reported success in prediction with the EPPS scales.

#### REFERENCES

- BUROS, O. K. *The fifth mental measurements yearbook*. Highland Park, New Jersey: Gryphon, 1959.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1954.
- GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1958.
- GEBHART, G. G., & HOYT, D. P. Personality needs of under- and overachieving freshmen. *J. appl. Psychol.*, 1958, **42**, 125-128.
- KRUG, R. E. Over- and underachievement and the Edwards Personal Preference Schedule. *J. appl. Psychol.*, 1959, **43**, 133-136.
- MCNEMAR, Q. At random: Sense and non-sense. *Amer. Psychologist*, 1960, **15**, 295-300.

(Received November 29, 1960)

# A NOTE ON RELATIONSHIPS BETWEEN STUDENT AND PARENT MINNESOTA TEACHER ATTITUDE INVENTORY SCORES

ROBERT E. CAMPBELL AND JOHN E. HORROCKS

*Ohio State University*

Influences contributing to the formation of attitudes toward teaching and pupil-teacher relationships are complex. They are also difficult to identify with any degree of comprehensiveness or accuracy. Riccio (1959), in a survey and study of the literature of teacher attitudes identified a number of related variables. The variables included personality traits, experiences, values, and, to some extent, academic aptitude. To the best of the writers' knowledge, an overlooked variable has been the role of parents in helping form the professional attitudes of those of their children who become teachers. The present study deals with possible relationships between professional attitudes of teachers-in-training and those of their parents. Three questions were considered: How similar are the attitudes of students and their parents? Does the student tend to identify attitudinally more with one parent than with the other? What is the relationship between the attitudes of a student's parents?

## PROCEDURE

The Minnesota Teacher Attitude Inventory (MTAI) (Cook, Leeds, & Callis, 1952) was administered to 127 Ohio State University College of Education students and their parents. The students were mainly sophomores enrolled in an educational psychology course. Both students and parents volunteered to participate in the study as a partial course requirement. The MTAI was administered to the students in a classroom session at the beginning of the course. A month later inventories and an explanatory letter were sent to parents. The letter stated that the investigators were interested in parental attitudes toward education. The parents were requested to answer their questionnaires separately and to refrain from showing their answers to each other. Returned answer sheets were carefully screened for misinterpretation of directions.

## RESULTS AND DISCUSSION

Table 1 shows the relationship between the MTAI scores of students and those of their parents. The data suggest three trends: the moderate positive (.45) agreement between parents' attitudes; the similarity of attitudes between students and their parents; and though tenuously, the students' attitudinal identification with the parent of the same sex. Daughters and mothers present a higher relationship (.37) than did daughters and fathers (.15).<sup>1</sup> Similarly the son-father correlation (.33) was higher than the son-mother correlation (.19). Although these differences were not statistically reliable, trends in these directions were not unexpected in view of similar results reported in studies made in other areas of parent-child identification. Father-son and mother-daughter identifications were noted by Gray and Klaus (1956) in their study of values, and by Henderson (1958) in his study of interests. Lazowick (1955), using an S-R approach, has also reported on the nature of identification.

Table 2 presents the means and standard deviations for the categories of groups used. Although the variance

<sup>1</sup> The difference between the correlations was not statistically significant at the .05 level of confidence.

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN STUDENTS' AND PARENTS' MTAI SCORES

Group	<i>r</i>	<i>N</i>
Students-fathers	.285*	127
Students-mothers	.334*	127
Mothers-fathers	.449*	127
Sons-fathers	.330*	39
Sons-mothers	.189**	39
Daughters-fathers	.149**	88
Daughters-mothers	.369*	88

\* Significant at .01 level of confidence.

\*\* Significant at .05 level.

## SUMMARY AND CONCLUSIONS

Similarities of teaching attitudes of 127 teachers-in-training and their parents were investigated by means of the Minnesota Teacher Attitude Inventory.

The results suggest the following conclusions: (a) parents are in relatively moderate agreement, (b) a fair positive relationship exists between the attitudes of students and those of their parents, (c) students tend to identify attitudinally with the parent of the same sex, (d) daughters are the most permissive in their concept of teacher-pupil relationships while sons are less so, and (e) mothers and fathers tend to be more authoritarian in their MTAI responses than do their children of either sex.

## REFERENCES

- COOK, W. W., LEEDS, C. H., & CALLIS, R. Minnesota Teacher Attitude Inventory manual. New York: Psychological Corporation, 1952.
- GRAY, S. W., & KLAUS, R. Assessment of parental identification. *Genet. psychol. Monogr.*, 1956, 54, 87-109.
- HENDERSON, H. L. The relationship between interests of fathers and sons and sons' identification with fathers: The relationship of the adolescent son's identification with his father to father-son interest similarity as measured by the Strong Vocational Interest Blank. Unpublished doctoral dissertation, Columbia University, 1958.
- LAZOWICK, L. M. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, 51, 175-203.
- RICCIO, A. C. The relationship of selected variables to attitudes toward teaching. Unpublished doctoral dissertation, Ohio State University, 1959.

(Received December 29, 1960)

TABLE 2

MEANS AND STANDARD DEVIATIONS OF MTAI SCORES FOR PARENTS AND STUDENTS

Group	<i>N</i>	<i>M</i>	<i>SD</i>
All students	127	30.4	28.4
Mothers	127	1.0	34.5
Fathers	127	-14.3	29.0
Daughters	88	35.4	28.2
Sons	39	18.9	25.2

is homogeneous for all groups, the means are significantly different. Since the MTAI may be interpreted on an authoritarianism-permissiveness continuum (high positive scores represent permissiveness in teacher-pupil relationships whereas low positive or negative scores present authoritarian attitudes) it is possible to rank daughters as most permissive with sons, mothers, and fathers following in that order. It would appear that an age as well as a sex variable may have been operating here.

## TEST ANXIETY AND THE INTELLECTUAL PERFORMANCE OF COLLEGE STUDENTS<sup>1</sup>

IRWIN G. SARASON<sup>2</sup>

*University of Washington*

A currently strong trend in the selection of college students is the reliance on objective tests of intelligence and aptitude. Very likely the problems associated with an increasing college age population and the heightened demands for admission to colleges are going to accelerate this trend. While the value of entrance examinations and objective measures of intellectual performance is obvious, a legitimate question would seem to be whether or not factors other than ability and aptitude affect the test performance of students. If it were the case, for example, that certain personality characteristics correlated negatively with intellectual performance measures, it would seem important to take this correlation into account in some way in evaluating applicants.

Although the problem of the possible confounding of personality with intellectual variables has been stated specifically in terms of admission to college, it is clear that this is but one aspect of the more general problem of the relationship between intelligence and personality. This is so because all students in most school systems undergo systematic intellectual assessment, and educational decisions frequently are made on the basis of this assessment.

<sup>1</sup>This research was supported by a grant from the Agnes H. Anderson Research Fund.

<sup>2</sup>The writer is indebted to Paul Horst and the staff of the Counseling and Testing Service of the University of Washington for their cooperation in making available these data.

The present research investigated the relationship between intellectual and personality measures for male and female freshman and sophomore college students. Thirteen intellectual measures were available for each student, of which six were high school grade point averages in a variety of fields, and seven were objective tests included in an entrance battery. Six personality variables were correlated with each of these 13 intellectual measures. Three of these six personality variables were measures of what might be called general anxiety, i.e., anxiety as expressed or experienced in a variety of situations. One other measure tapped the student's feelings of hostility in a variety of situations and another measure was designed to assess the subject's level of defensiveness. The sixth personality variable studied, and the one which it was felt would most strongly relate to intellectual performance, was test anxiety, i.e., anxiety experienced specifically in testing situations. The prediction of significant relationships between test anxiety and intellectual performance was based on the following formulation.

Anxiety responses are aroused by personally threatening conditions in the environment. These responses interfere with ongoing task-relevant activity and lead to a lowering of performance level. Although the exact nature of the interfering anxiety responses is not known, it seems likely that they include both heightened

physiological activity (e.g., heart rate, palmar sweating) and self-deprecating ruminations (e.g., "I can't pass this test," "I wonder how the other students are doing"). While these anxiety responses are assumed to be aroused under stressful conditions, their evocation would not necessarily be expected under neutral conditions (Sarason, 1960).

Within this view of anxiety, anxiety scales or questionnaires may be considered to measure the likelihood of emission of anxiety responses under conditions of personal threat. Furthermore the closer the content of the anxiety scale is to the situation in which the subject's performance is measured the better predictor of reaction to threat will the anxiety scale be. Since the test anxiety scale used in this study contains items dealing with the subject's reactions to testing situations, it was expected that it would most meaningfully relate to performance. Because of the importance of college aptitude tests to students it was predicted that test anxiety would correlate negatively with the seven entrance examination tests. The correlations of test anxiety with the six high school grade point averages were expected to be of a lower order than those with the entrance test battery. This latter prediction was made because of the possibility that while highly anxious students might not be able to prepare for and cope adequately with entrance and aptitude tests, they well might be able to reduce their anxiety levels over a semester or school year in course work.

#### METHOD

*Subjects.* The subjects were 326 male and 412 female students enrolled in introductory psychology and introductory sociology courses at the University of Washington. The Autobiographical Survey containing

the personality measures (Sarason, 1958b) was administered during regular class sessions. The aptitude measures were obtained prior to the subjects' entering the University of Washington. Thus the personality and intellectual measures were not obtained contemporaneously.

*Materials.* The form of the Autobiographical Survey used in this research consists of six true-false scales: Test Anxiety (TA), 17 items; General Anxiety (GA), 18 items; Lack of Protection (LP), 27 items; Hostility (H), 22 items; Need for Achievement (NA), 31 items; Defensiveness (D), 24 items. Descriptions of the scales have been presented previously (Sarason 1958b).<sup>3</sup>

The 13 intellectual measures employed are routinely gathered for entering University of Washington students. The measures are:

1. High school English average
2. High school mathematics average
3. High school foreign language average
4. High school social studies average
5. High school natural science average
6. High school electives average
7. Verbal score on the Guilford-Zimmerman Aptitude Survey
8. Mechanical Knowledge score on the Guilford-Zimmerman Aptitude Survey
9. Cooperative English Usage score
10. Cooperative English Spelling score
11. Cooperative Mathematics I score
12. Cooperative Social Studies II score
13. ACE (1948) Q score

#### RESULTS

The results for the male and female subjects will be presented first in terms of the correlations between the intellectual and personality measures and then in terms of the intercorrelations of the Autobiographical Survey scores.

Table 1 contains the correlations between the 6 personality and 13 intellectual variables for the 326 male subjects. The table also contains the means for the 13 intellectual measures. Although GA correlated negatively and significantly with 4 intellectual measures and NA correlated negatively and significantly with 5 of these

<sup>3</sup> Copies of the Autobiographical Survey may be obtained from the author.

TABLE 1  
CORRELATIONS OF SIX AUTOBIOGRAPHICAL  
SURVEY SCORES WITH THIRTEEN  
INTELLECTUAL MEASURES FOR  
326 MALES

Measure	TA	GA	LP	H	NA	D	Mean
High school English average	-09	-04	.00	-06	-05	.06	2.82
High school mathematics average	-18	-08	-01	-03	-09	.08	2.61
Foreign language average	-09	.03	.09	-03	-07	.00	2.57
Social science average	-17	-04	-01	-08	-01	.07	2.95
Natural science average	-15	-07	.02	-07	-06	.07	2.78
Elective average	-14	-07	-04	-12	-09	.07	3.06
G-Z Verbal	-23	-03	.00	.05	-18	.00	27.18
G-Z Mechanical Knowledge	-13	-11	-07	.00	-16	.03	21.85
Cooperative English Usage	-27	-12	-02	-06	-14	.11	86.61
Cooperative English Spelling	-18	-05	-04	.00	-11	.03	16.17
Cooperative Mathematics I	-13	-10	-03	.00	.00	.11	23.71
Cooperative Social Studies II	-28	-11	-01	.06	-13	.06	18.45
ACE Q score (1948)	-30	-17	-06	-01	-06	.10	46.98

Note.— $r = .133$  at .05 level;  $r = .148$  at .01 level.

measures, it is clear that TA shows the most consistent set of correlations. For 11 of the 13 intellectual measures there were significant negative correlations with TA.

A comparable set of correlations for 412 females (Table 2) also shows TA to correlate negatively and significantly with 11 of the 13 intellectual measures, although the same 11 variables were not significant for both males and females. The only other personality characteristic found to be related to intelligence was H which correlated positively with four intellectual measures.

For both the male and female groups, comparisons were made of the correlations obtained between TA and high school grade point averages and TA and entrance tests. For the purposes of this comparison the six TA

grade point averages were compared with six TA entrance test scores. Not considered were the male and female TA-Mechanical Knowledge correlations. This omission was based on the great differences between male and female means on this test. The comparisons made indicated stronger negative correlations between TA and the six entrance tests than between TA and the six high school grade point averages. This result was significant for both males and females ( $p < .05$  and  $p < .02$ , respectively).

Intercorrelations between the six Autobiographical Survey measures and the means for these measures are presented in Tables 3 and 4, for males and females, respectively. In general the patterns of intercorrelations for males and females are quite comparable. For both sexes the D measure is

TABLE 2  
CORRELATIONS OF SIX AUTOBIOGRAPHICAL  
SURVEY SCORES WITH THIRTEEN  
INTELLECTUAL MEASURES FOR  
412 FEMALES

Measure	TA	GA	LP	H	NA	D	Mean
High school English average	-12	.02	.01	.02	.00	-.02	3.30
High school mathematics average	-12	.03	.06	.01	-.01	-.07	2.81
Foreign language average	-14	.10	.05	.07	-.05	-.07	3.08
Social science average	-13	.02	.03	.05	-.02	-.04	3.24
Natural science average	-19	-.01	.05	.06	-.04	.00	3.06
Elective average	-.06	.00	.01	-.01	.05	-.01	3.33
G-Z Verbal	-.22	.00	.07	.13	-.07	-.03	27.55
G-Z Mechanical Knowledge	-.08	-.09	-.04	.10	-.05	-.03	4.51
Cooperative English Usage	-25	.00	.06	.09	-.03	-.02	101.78
Cooperative English Spelling	-16	.03	.00	.11	-.04	-.03	20.47
Cooperative Mathematics I	-21	-.04	.02	.09	-.01	.00	17.15
Cooperative Social Studies II	-29	-.05	.04	.11	-.08	-.01	16.43
ACE Q score (1948)	-18	-.04	.01	.07	.06	-.01	43.88

Note.— $r = .098$  at .05 level;  $r = .128$  at .01 level.

TABLE 3  
INTERCORRELATIONS OF SIX AUTO-BIOGRAPHICAL SURVEY SCORES FOR 326 MALES

Scores	TA	GA	LP	H	NA	D	Mean
TA		53	35	13	34	-50	4.97
GA			58	23	34	-64	5.57
LP				33	37	-56	6.32
H					20	-40	8.86
NA						-35	11.57
D							12.50

Note.— $r = .113$  at .05 level;  $r = .148$  at .01 level.

TABLE 4  
INTERCORRELATIONS OF SIX AUTO-BIOGRAPHICAL SURVEY SCORES FOR 412 FEMALES

Scores	TA	GA	LP	H	NA	D	Mean
TA		48	35	08	35	-45	5.64
GA			54	26	31	-60	6.40
LP				33	24	-55	7.08
H					14	-44	6.96
NA						-31	10.97
D							13.42

Note.— $r = .098$  at .05 level;  $r = .128$  at .01 level.

negatively related to each of the other Autobiographical Survey measures.

### DISCUSSION

The present findings demonstrate that for both men and women there are significant negative relationships between anxiety reported by subjects as being experienced in testing situations (TA) and a variety of measures of intellectual performance. These results also confirm the findings of a previous study (Sarason, 1959) which showed (a) TA to be negatively related to many of the same intellectual variables with which the present study was concerned and (b) other personality measures to be unrelated to intellectual performance. In the present research none of the personality scales, except TA, related significantly to

performance in any consistent manner for both male and female college students. Thus, support was given for the hypothesis that the more specific the measure of personality (in this case, anxiety) is to the situation being studied, the more consistently will the personality measure relate to performance (Sarason, 1960). It was hypothesized also that TA scores would be more negatively related to aptitude test measures than to the six high school grade point averages. The present study suggests that this well may be the case. The Need for Achievement scale might have been expected to relate negatively to intellectual performance since its items appear to reflect strong need and anxiety to achieve. There was an indication that this might be true for male students but not for female students.

Two problems then seem to be posed by the present findings of consistent negative relationships between TA and 13 intellectual measures and also of greater strength of those negative relationships involving aptitude tests as compared with high school grade point averages. First, what factors contribute to the negative relationships obtained? Second, why is test anxiety less of a detrimental factor for course grades than for aptitude tests?

With respect to the first question, it would seem prudent at present to interpret the present results as simply indications of poorer intellectual performance for high than for low test anxious students. If this is due to intellectual differences among subjects differing in TA scores, then it may be the case that high test anxiety scores simply reflect realistic concern over intellectual ability. If, however, students differing in test anxiety do not differ in intellectual ability then it would be worthwhile to determine

which psychological aspects of testing and academic situations arouse anxiety responses in high TA scorers and whether or not the conditions of either the testing situation or of anxious students can be modified in such a way as to raise anxious students' performance level. Results of experimental studies indicate that under neutral or reassuring conditions high and low anxious subjects do not differ in their performance levels but that under stress they do differ in the same direction as obtained in the present research (Sarason, 1958a; Sarason & Palola, 1960).

In view of this evidence it seems possible that the answer to the second question posed, that of the explanation of the stronger correlations with TA of aptitude test scores than of grade point averages, may be in the opportunity of the student in class to be reassured and to reassure himself. In the course of a quarter or a semester the anxious student can come to know his instructors and peers, as well as to increase his familiarity with and to overlearn the material to be learned. These possibilities are largely absent in group aptitude testing. Greater emphasis in future research on the interpersonal aspects of learning and performance situations should make clearer the implications of these findings.

Theoretically, the present study has supported the interfering response interpretation of anxiety discussed earlier. These results are consistent with a conception of the TA as an index of proneness to performance disruption under personally threatening conditions. Practically, these findings strongly suggest the need to take account of the test anxiety factor in interpreting students' intellectual performance.

One factor of importance in the

present study concerns that of the population of subjects used. While TA correlated consistently with the intellectual measures used, in most cases the magnitude of the  $r$ 's was not great. It seems possible that the restriction in range of intelligence inherent in a college sample may have worked against obtaining even more significant  $r$ 's than were found in the present research. Future research with high school level subjects and a greater range in intellectual ability should, if this is correct, lead to larger  $r$ 's than were obtained in this study.<sup>4</sup>

The intercorrelations of the Autobiographical Survey scales show essentially the same patterns for the women as for the men. Interestingly the present intercorrelations for college students are quite similar to the intercorrelations previously reported for hospitalized psychiatric patients (Sarason, 1958b). The only sizable differences in the correlations for the two subject populations are the  $r$ 's involving the Hostility scale. For the patient population this scale correlated much more strongly with the other Autobiographical Survey measures than was true in the case of college students.

#### SUMMARY

The present study correlated scores on six personality scales of the Autobiographical Survey with scores on seven aptitude test scores and six high school grade point averages. The subjects were 326 male and 412 female college students.

The results demonstrated that of the six personality scales only the Test Anxiety scale (TA) related consistently to the 13 intellectual measures. For both males and females, TA

<sup>4</sup> A recently completed study conducted by the writer with high school students as subjects provides support for this expectation.

correlated negatively and significantly with 11 of the 13 intellectual measures. There were indications, however, that the negative correlations with TA were stronger for aptitude test scores than for grade point averages. These results were interpreted as supporting the hypotheses that: (a) the closer the content of anxiety scales' items is to the situation in which subjects are to perform the more predictive will the items be, (b) TA scores reflect the proneness of subjects to emit interfering anxiety responses under conditions of threat in a testing situation.

Comparisons of the intercorrelation matrices of the six Autobiographical Survey scales for males and females showed quite similar patterns. The Autobiographical Survey intercorrelations for male and female college students were found to be similar to

those previously reported for a neuro-psychiatric population.

#### REFERENCES

- SARASON, I. G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *J. exp. Psychol.*, 1958, **56**, 472-477.(a)
- SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339-344.(b)
- SARASON, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1959, **59**, 272-275.
- SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403-415.
- SARASON, I. G., & PALOLA, E. G. The relationships of test and general anxiety, difficulty of task, and experimental instruction to performance. *J. exp. Psychol.*, 1960, **59**, 185-191.

(Received January 23, 1961)

# EFFECTS OF MEMORY, EVALUATIVE, AND CREATIVE READING SETS ON TEST PERFORMANCE

E. PAUL TORRANCE AND JUDSON A. HARMON

*University of Minnesota*

For several years the senior author (Torrance, 1950, 1951) has been experimenting with a variety of procedures which might increase the likelihood that students in mental hygiene courses will make use of scientifically developed principles of mental health in their work. Several recent experiments (Torrance, 1960) have been motivated by the belief that the failure of educators to make more effective use of such principles results from the fact that they have not been educated to think creatively or imaginatively about course materials. It soon became evident in this experimentation that many, perhaps most, students did not have the expectation that research results or scientifically developed principles *could* be applied imaginatively in solving many of their personal and professional problems. At best they were looking for specific prescriptions.

The rather shocking realization that students enrolled in a graduate course in mental hygiene did not even expect to be able to use the knowledge they were acquiring led the senior author to try to develop ways of establishing appropriate "sets." In one study (Torrance, 1960) random halves of two classes were asked to read research articles creatively or critically. Those who read and reported on the articles creatively were later more successful in developing a new idea of their own and in responding to a test of creative applications. These results stimulated the present study, an extension of "sets" to textbook readings.

The effect of appropriate prepara-

tory sets in facilitating learning has been well-established (Johnson, 1955, p. 77). Usually a person's activities are integrated and he does not react to every event that stimulates his sense organs. He is selective not only of what he responds to but also of what is produced and of the mental operation used in responding.

In designing the present experiment, the authors have been guided by Guilford's (1959) model of human intellect. Particular attention has been paid to his five mental operations: cognition, memory, convergent thinking, divergent thinking, and evaluation. By "cognition," Guilford means the operations involved in recognizing facts, discovery, and rediscovery. Memory involves the retention of what has been cognized. The two kinds of productive thinking, convergent and divergent, include the generation of new information from known information and remembered information. Convergent thinking leads to a conventional solution; divergent thinking leads to new and untested solutions. Evaluation refers to decisions as to the goodness, suitability, or adequacy of what we know, remember, or produce through convergent or divergent thinking. In this experiment a deliberate attempt was made to establish sets which would result in utilization of these operations.

## METHOD

### *Subjects*

The subjects were 115 students in a graduate level course in Personality Development and Mental Hygiene at the Uni-

versity of Minnesota during the summer of 1960. Seventy-six were males and 39 were females. Most subjects had had professional experience in one or more phases of education; 5 were public health nurses and 10 had had no professional experience. The average number of years of experience was 6.6 years and the range was from 0 to 30 years.

Subjects were assigned alphabetically to one or another of three groups (1, 2, or 3). The group number signified the reading set to be followed during a period of one week. This method of assignment to groups resulted in no biases due to sex, professional field, or number of years of experience. Some members of the original sample canceled the course or otherwise failed to participate in certain phases of the study, making the size of the three groups slightly unequal.

### Description of the Reading Sets

The instructions of the memory, evaluation, and creative application sets were given to the subjects as follows:

*Set A—Memory.* When you read, it is important that you read with a set or intent to remember everything you read. In reading your assignments for the next three days, practice using this set. Try to make your mind act as a "sponge" and "soak up" as much as possible of whatever you read. Use whatever devices you like for trying to remember what you read. It may take some practice before you are really successful in assuming this set, but do not be discouraged. By the third day, you should find it easy to assume this set.

*Set B—Evaluation.* When you read, it is important that you think about what you are reading and evaluate critically the content. In reading your assignments for the next three days, practice using this set. As you read, ask: "Is what the author is saying really true? Does what he is saying agree with what I have experienced? Does what he is saying agree with what is known through scientific research?" Try to find defects in the reasoning of the author, in his factual presentation, and the like. Do not be concerned about errors in grammar, punctuation, sentence structure, and the like. It may take some practice . . . (etc., same as above).

*Set C—Creative Application.* When you read, it is important that you think about the many uses of the information which you are reading. It is especially important that you think of the various ways in which the information could be used in

your personal and professional life. In reading, do not just ask, "What is the author saying?" Also ask, "How can I use what the author is saying?" Do not stop with just one use. Think of as many uses as you can of the important ideas presented. Jot down some of these uses for future reference. It may take some practice . . . (etc.).

### Experimental Procedure

At the beginning of the first experimental week, subjects received reading-set instructions in accord with the group to which they had been assigned. Each subject received only one set during any one week. Thus, during the first week subjects in Group 1 received Set A; those in Group 2, Set B; those in Group 3, Set C; and so on—sets being shifted from group to group over the 3 weeks.

Each week each subject was provided an Evaluation of Reading Set Form upon which he was asked to report the degree to which he was able to maintain the assigned set, the degree of difficulty experienced in using that set, the number of minutes spent in completing the reading assignment for the last day of the week, and semantic reactions to the experience. Osgood's (1952) semantic differential technique was used in obtaining the latter. The following polar-adjective pairs were used with seven-point scales: good-bad, pleasurable-painful, labored-easy, wise-foolish, active-passive, fast-slow, cautious-rash, new-old, and interesting-boring.

Each Friday subjects were administered a 20-minute test which covered the reading assignment for that day. This test included items which were defined as: cognitive (multiple-choice), memory (fill-in-the-blanks), divergent thinking (extensions from given facts, uses of facts, questions, hypotheses, etc.), and evaluation (decision making situations and justifications of decisions). It was at this time that the Evaluation of Reading Set Forms were collected.

Part-scores based on performance on these subtests were determined for all subjects and were reported to each subject at the beginning of the next week.

The following are samples of items presumed to involve cognition:

According to a recent study by two New York scientists involving tests of physical fitness, which of the following findings was obtained?

—American children superior to European children.

- European children superior to American children.
- American children superior to Russian children.
- Russian children superior to American children.
- American children superior to Australian children.

Which of the following is *least likely* to contribute to the prevention of paranoid behavior?

- Training in straight thinking.
- Fostering attitudes of good sportsmanship.
- Finding available objects to blame.
- Developing self-ennobling attitudes toward sex.
- Facilitating social interaction.

The following are examples of the memory items:

According to Wang, the primary factor in personality maladjustment is either \_\_\_\_\_ or \_\_\_\_\_ of the child's need for parental affection and sympathetic handling.

The experimenter who used a mechanical mother in laboratory studies at the University of Wisconsin is \_\_\_\_\_.

The following illustrate the creative-application problems:

If hunger needs are not satisfied during childhood, how is this likely to affect personality development? Assume that the child has been given enough food for subsistence but rarely enough to satisfy hunger. What kind of adult personality would you expect?

Research has shown that lower socio-economic individuals have limited ability to delay gratification. List as many ways as you can think of for using this information in educational situations to promote healthy personality adjustment.

The following is a sample of the evaluative problems:

Assume that you are a member of the controlling board of the Minneapolis General Hospital which is considering the establishment of a children's psychiatric unit for mentally disturbed children. The treatment program will be quite similar to that described in the text for the Children's Psychiatric Unit at the University of Michigan and it is estimated that the cost of treatment per child will be \$6,500. In deciding this issue, how would you vote?

— Yes — No  
How would you justify your vote? Cite as many sound reasons as you can.

### *Analysis of Data*

Analyses of variance were computed, comparing the effects of the reading sets of each of the different kinds of test performance (cognitive, memory, divergent thinking, and evaluation) for each of the 3 weeks. The evaluative-type item, unfortunately, was used in only one of the tests. Analyses of variance were computed for success in maintaining set, difficulty in maintaining set, number of minutes spent in completing the reading assignment, and each of the nine semantic differential scales for the three sets. Analyses of variance were also computed on Test Scores  $\times$  Weeks and for Weeks  $\times$  Groups in order to determine relative performance by weeks and by groups.

### RESULTS

The data comparing scores on the memory-type items are presented in Table 1. It will be noted that only during the third week were the differential effects of the three sets statistically significant at the .05 level of significance or better, the memory set producing the highest score. The predicted trend occurred during the second week but was not statistically significant.

Table 2 presents the means and analysis of variance data for scores on the Creative-Applications or divergent-thinking items. It will be noted that the differential effects are rather strong for each of the 3 weeks and that in all cases those working under the creative set achieved the highest mean score. Those using the evaluative set consistently achieved the next highest mean score. On the third test all three groups tended to achieve higher scores than during the earlier weeks. This is probably due to the accumulative effects of the course which emphasized creative thinking concerning the course content.

The data concerning performance on the evaluative (decision making) items for the second week are shown in Table 3. The differential effects are statistically significant, with the eval-

TABLE 1  
COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE IN  
MEMORY TYPE TESTS (COMPLETION) FOR EACH OF THREE WEEKS

Week	Reading set	N	M	Between SS	Within SS	F ratio
First	Memory	40	6.04	3.9259	195.7898	2.226
	Evaluative	38	6.24			
	Creative	36	5.78			
Second	Memory	33	7.32	9.8958	420.2684	2.543
	Evaluative	41	6.65			
	Creative	37	7.20			
Third	Memory	38	8.30	12.7482	405.2544	3.429*
	Evaluative	34	8.15			
	Creative	40	7.54			

\* p < .05.

TABLE 2  
COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE IN  
CREATIVE-APPLICATION TYPES OF TESTS FOR EACH OF THREE WEEKS

Week	Reading set	N	M	Between SS	Within SS	F ratio
First	Memory	40	4.25	72.4602	643.6542	12.496**
	Evaluative	38	5.43			
	Creative	36	6.18			
Second	Memory	33	4.45	262.6431	372.4977	76.1493**
	Evaluative	41	4.94			
	Creative	37	7.96			
Third	Memory	38	6.88	66.2446	524.4794	13.7613**
	Evaluative	34	8.37			
	Creative	40	8.60			

\*\* p < .01.

TABLE 3  
COMPARISON OF EFFECTS OF THREE  
READING SETS ON PERFORMANCE IN  
EVALUATIVE (DECISION MAKING)  
TYPE OF TEST FOR SECOND WEEK

Reading set	N	M	Be-tween SS	Within SS	F ratio
Memory	33	5.88	—	—	—
Evaluative	41	7.34	39.1292	368.4517	11.469**
Creative	37	6.70	—	—	—

\*\* p < .01.

uative, creative, and memory sets ranking in that order. Although this type of item was used at other times during the course, the only test in the

experiment containing such an item was the one for the second week.

Table 4 presents the data for the cognitive items for each of the three tests. It had been anticipated that subjects using the memory set would perform best on these items. This occurred only during the second week. The differential effects were also significant for the last test, but subjects working under the evaluative set achieved the highest mean score. It will also be noted that subjects using the evaluative set also had a slight edge during the first week but the differential effects were not statistically significant.

TABLE 4

COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE ON COGNITIVE TYPE OF TESTS (MULTIPLE-CHOICE) FOR EACH OF THREE WEEKS

Week	Reading set	N	M	Between SS	Within SS	F ratio
First	Memory	40	6.21	1.9062	236.9290	0.893
	Evaluative	38	6.52			
	Creative	36	6.33			
Second	Memory	33	4.03	10.3007	200.9297	5.537**
	Evaluative	41	3.32			
	Creative	37	3.43			
Third	Memory	38	4.68	7.3804	211.8738	3.797*
	Evaluative	34	5.06			
	Creative	40	4.42			

\* $p < .05$ .

\*\* $p < .01$ .

### Test Performance by Group and by Week

Analyses of variance were also performed to study differential test performance by group and by week. It was found that the effects for groups are not statistically significant. The effects for weeks are statistically significant, total mean scores being higher for the third week than for the first two weeks. Thus, there is no evidence of the superiority of any one sequence of sets over the others.

### Other Reactions to Assigned Sets

Failure to maintain an assigned set is obviously an important consideration in determining the effectiveness of a set in producing the desired results. Two questions on the Evaluation of Reading Set Form were designed to supply information concerning this problem. Responses concerning the completeness with which the set was maintained ranged from "not at all" (Point 1) to "almost all of the time" (Point 7). Responses concerning difficulty in learning to use the set ranged from "impossible" (Point 6) to "extremely easy" (Point 6). Weights were assigned arbitrarily.

Means and F ratios are presented in Table 5 for the variables: success in using set, difficulty in using set, number of minutes spent in reading assignment for day, and for each of the semantic scales.

From Table 5 it will be noted that the sets had significant differential effects on both success and difficulty in using set but not on time spent on reading assignment. The memory set seems to have been the most successfully and easily maintained and the creative set seems to have been the least successfully and easily maintained. Subjects estimated that they maintained the memory set on the average about two-thirds of the time, finding it between "fairly difficult" and "fairly easy" to maintain. They reported that they were successful in using the creative set less than half of the time.

From Table 5 it will also be noted that on only four of the nine polar-adjective pairs are reactions significantly different among the three conditions of reading set. Those pairs are: fast-slow, cautious-rash, new-old, and interesting-boring. Subjects using the memory set tended towards fast, cautious, old, and boring. Those using the

TABLE 5

COMPARISON OF EFFECTS OF THREE READING SETS ON SUCCESS IN USING SET,  
DIFFICULTY IN USING SET, TIME SPENT ON READING ASSIGNMENT, AND  
SEMANTIC REACTION FOR THREE WEEKS COMBINED

Variable	Memory		Evaluative		Creative		F ratio
	N	M	N	M	N	M	
Success in using set	106	5.18	109	4.06	105	3.79	63.35**
Difficulty in keeping assigned set	106	4.54	110	4.10	105	3.97	26.72**
Number of minutes spent on reading assignment	104	112.27	110	114.52	105	112.00	0.23
Good-Bad	105	2.47	109	2.92	102	2.59	0.59
Pleasurable-Painful	103	3.08	107	3.16	103	3.01	0.87
Labored-Easy	103	3.96	107	3.64	103	3.73	2.44
Wise-Foolish	103	2.45	109	2.37	102	2.36	0.30
Active-Passive	103	2.64	107	2.67	102	2.48	1.46
Fast-Slow	102	3.96	107	4.38	103	4.33	5.02**
Cautious-Rash	102	2.86	107	2.97	102	3.13	3.27*
New-Old	103	4.55	107	3.64	103	2.88	18.38**
Interesting-Boring	103	2.76	107	2.25	102	2.13	17.11**

\* $p < .05$ .

\*\* $p < .01$ .

evaluative set tended towards slow and held to a middle position on the other pairs. Under the creative set the direction was towards rash, new, and interesting.

In order to determine whether or not there are differential effects with the passage of time, a comparison was made by weeks on the variables listed in Table 5. Only four of the variables gave rise to changes with time as data of analyses of variance showed. During the second week subjects tended to report greater success in maintaining the sets than during the other weeks. They also tended to spend somewhat more time on the reading assignment during this particular week. Finally, there was a consistent change towards fast and boring over the 3 weeks. Of these four variables the time spent on the reading assignment was significant at  $.01 < F < .05$ . The other three were significant at  $<.01$  level.

#### DISCUSSION

From the data which have been presented it seems clear that instructing mature students to assume various reading sets will have differential effects on the kind of goals achieved. This would seem to be especially true if creative thinking and evaluative thinking are desired as outcomes. Although the students in the present study were able to maintain the creative and evaluative sets only about half of the time, attempts to maintain these sets had measurable effects on performance on tests covering the reading assignment.

Since most examinations given to large classes are of the cognitive (multiple-choice) type, it is worthwhile to note that the memory set does not uniformly result in the best performance on this type of test. In fact, in only one of the 3 weeks does the memory set result in superiority on the cognitive type of test. The evaluative set

tended to result in better mean scores during the other 2 weeks.

Because of the differential effects of the three sets on the four types of tests, one might also infer that instructors need to develop tests designed to measure different kinds of achievement, depending upon course objectives. If instructors expect their students to think imaginatively and evaluatively about what they learn, it is extremely doubtful that such goals are adequately assessed by the traditional multiple-choice and completion items.

The data indicate that students are more successful in maintaining the memory set and find it easier to maintain this set than the creative and evaluative sets. This probably results from the fact that they have had much more practice with this set than with the others. This hypothesis receives some support from reactions to the semantic differential. For example, subjects under the memory set tend to rate the experience in the direction of cautious, old, and boring. It is likely that, with practice, students will find it easier to maintain creative and evaluative sets and that better results will be achieved than those reported herein.

Finally, in interpreting the results, it should be recognized that the subjects are graduate students and that almost all of them have had rather extensive experience in various types of educational work.

#### SUMMARY

This experiment was designed to test the differential effects of memory, evaluative, and creative "sets" in reading assignments in a graduate level course on Personality Development and Mental Hygiene. The 115 subjects were arranged alphabetically

and assigned to one of three groups. Each group was assigned in turn each of the three reading sets to be used in the assigned readings for a week. At the end of each week, a 20-minute test containing cognitive, memory, creative thinking, and evaluative items was administered. Subjects also estimated their degree of success in maintaining the set, difficulty in maintaining it, the number of minutes spent on the Friday's reading assignment, and their semantic reactions to nine polar-adjective pairs.

Analyses of variance indicated that the three sets produced differential effects on most of the tests for all 3 weeks. Those operating under the creative set, achieved the highest mean each week on the creative applications. Those operating under the evaluative set achieved the highest mean on the evaluative or decision making problem when such a problem was given (second week only). Those operating under the memory set achieved the highest mean on the memory tests (completion) only during the third week. The differential effects on the memory items were not statistically significant during the other weeks. Differential effects on the cognitive items were not consistent. Those operating under the memory set achieved the highest mean during the second week and those under the evaluative set led during the third week.

Subjects reported that they found it easier to maintain the memory set and rated themselves as more successful in maintaining this set than the other two. There were no differential effects on the estimated length of time spent on the reading assignment. On the semantic differential, subjects under the memory set tended towards fast, cautious, old, and boring. Those

under the evaluative set tended towards slow and those under the creative set, towards rash, new, and interesting.

During the second week subjects generally reported greater success in maintaining sets than during the first and third weeks. On the semantic evaluations, there was a consistent and significant drift towards fast and boring. The three original groups did not differ significantly, giving no evidence of the superiority of any one sequence of sets over the others.

#### REFERENCES

- GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist*, 1959, 14, 469-479.

- JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harpers, 1955.
- OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, 49, 192-237.
- TORRANCE, E. P. The phenomenon of resistance in learning. *J. abnorm. soc. Psychol.*, 1950, 45, 592-597.
- TORRANCE, E. P. Getting mental-hygiene practices into action through a college class. *Ment. Hyg., NY*, 1951, 35, 88-95.
- TORRANCE, E. P. Effects of induced evaluative sets on the development of new ideas among graduate students. In E. P. Torrance (Ed.), *Creativity: Second conference on gifted children*. Minneapolis: University of Minnesota, Center for Continuation Study, 1960.

(Received February 8, 1961)

# THE EFFECTS OF VISUAL DISCRIMINATION PRETRAINING WITH WORD AND LETTER STIMULI ON LEARNING TO READ A WORD LIST IN KINDERGARTEN CHILDREN

SIEGMAR MUEHL<sup>1</sup>

*Iowa Child Welfare Research Station*

Ability to discriminate among word forms is a prerequisite to associating these same stimuli with appropriate word names.

An earlier study by the author (Muehl, 1960) using kindergarten subjects, showed that visual discrimination pretraining with the same words to be read in a subsequent word list facilitated learning the word list when compared to discrimination pretraining with different words or pretraining with geometric forms. These results suggested the possibility that specific observing responses were learned to the stimuli in the same-word pretraining; and, that these responses transferred positively to discriminating among the words in the vocabulary task.

The purpose of the present experimental study was to determine, more specifically, the stimuli relevant to visual discrimination among word forms for beginning readers.

One assumption tested is that the child learns to discriminate among words on the basis of the shape or form of the whole word. A second assumption tested is that the child discriminates among words on the basis

of attending to some part of the word, such as a particular letter, part of a letter, or letter grouping.

The assumption that word shape is the basis for visual discrimination among words, although lacking objective confirmation (Davidson, 1931), appears to be supported by educators, judging by the widespread practice of the so-called "whole-word" method of introducing reading in the first grade. According to Gray (1956, pp. 83-84), this method is based on the assumption that, "...individual words are the basic units of thought and recognition," or, that, "...each word has a characteristic form by which it can be remembered." Although words are the basic meaning units in our language, it does not follow they are also the basic units of visual recognition.

The assumption that the beginning reader discriminates among words on the basis of letters finds support in two different areas of educational research. Comparing intralist errors in word recognition tasks for children ranging from 4 to 6 years in age, Gates and Boeker (1923), Meek (1925), and Muehl (1960) found that letter similarities were the apparent source of confusion among words. Using a camera to record eye movements, Buswell (1922) showed that beginning readers made many eye fixations associated with details within a word in the early reading stages.

The results of psychological re-

<sup>1</sup>The author wishes to thank Buford W. Garner, Superintendent of Schools, Iowa City; Marjorie Clatterbaugh, Donald Tvedt, and John Moore, principals; and Nancy Cooker, Audrey Nelson, Nellie Reiger, and Carol Ripp, kindergarten teachers whose cooperation made this research possible.

search bearing on the related question of part vs. whole discrimination are inconclusive. A review of this research with children, reported by Gibson and Olum (1960), showed that the research results were conflicting and appeared to depend on the stimulus material used. However, none of the studies reviewed had used word stimuli.

## METHOD

### *Pretraining Task*

To test these alternative assumptions, three groups of kindergarten children were differentiated by the kinds of visual discrimination or matching pretraining each received prior to learning to read a word list. The pretraining paradigm is shown in Table 1. The nonsense word sets for Groups RR and IR were designed so that either word shape (stimulus configuration determined by all the letters) or letter differences among the three words in each set could serve as relevant stimuli for discriminating among the words. LD pretraining was included to control for the effect of learning to discriminate among the relevant letters, as such, on later word list performance.

Each of the pretraining sets (RR, IR, and

LD) was presented in four orders. Two of these orders presented consecutively constituted a two-trial block. The words and letters appearing in the sets were typed in lower case with a primary-style typewriter. The word stimuli were approximately 11 mm. in length and 5 mm. in height to include the tallest letters. These word and letter stimuli were mounted in the center of  $2\frac{1}{2}'' \times 2\frac{3}{4}''$  white cards. Accompanying each two-trial block of stimulus items in the three sets were two,  $2\frac{3}{4}'' \times 8\frac{1}{2}''$ , multiple-choice response cards. The three words, or three letters, appearing on the stimulus cards in each set also appeared on these response cards, in two different position orders. The apparatus for presenting the stimulus and response cards has been described in detail elsewhere (Muehl, 1960). Briefly, it consisted of a tray for presenting the stimulus cards, a hinged window below the tray for exposing the response card, and three buttons located on a panel immediately in front of each of the three words appearing on the exposed response card. If the subject selected the correct response word and pressed the button under it a door chime would ring. The experimenter controlled the appropriate button-chime connection by means of a selector switch from behind the apparatus.

All subjects received eight matching trials. A presentation consisted of the experimenter's placing a stimulus card on the

TABLE 1  
STIMULUS-RESPONSE PARADIGM TO ILLUSTRATE EXPERIMENTAL DESIGN

Visual Discrimination Pretraining		Reading Task	
Stimuli	Responses	Stimuli	Responses
Group RR <sup>a</sup>		All Groups	
feu		feu	
geu		geu	
reu		reu	
Group IR <sup>b</sup>	matching		
fjd			blue
gjd			red
rjd			white
Group LD <sup>c</sup>			
f			
g			
r			

LD

LD-O

same as above

same as above, plus orienting instructions

<sup>a</sup> Relevant shape-relevant letter.

<sup>b</sup> Irrelevant shape-relevant letter.

<sup>c</sup> Letter discrimination.

card-holder tray for approximately 2 seconds. The subject was told to look carefully at the card. The experimenter then removed the card and the subject was instructed to open the response card window, find the same stimulus item on the response card that had appeared on the stimulus card, and then press the button directly under it to find if his choice was correct. If the subject pushed the incorrect button on any one of the first three presentations, the experimenter showed the stimulus item again and had the subject find the correct response item and ring the bell. On subsequent presentations, if the subject responded incorrectly or if the subject did not respond in approximately 5 seconds, the experimenter closed the response card window and presented the next stimulus card. The experimenter alternated response cards with each two-trial block. No rewards were given in pretraining other than provided by the chime, and verbal encouragement from the experimenter.

### Reading Task

The reading or transfer task was designed to make both word-shape and letter differences among the Group RR pretraining words relevant stimuli for discriminating among the reading task words. For Group IR pretraining words, the reading task was designed to make letter differences alone relevant to discriminating among the reading task words. This reading task paradigm is shown in Table 1.

Set RR nonsense words were paired with  $\frac{5}{8}'' \times \frac{7}{8}''$  color patches. The typed words and color patches were mounted on  $3\frac{1}{2}'' \times 6\frac{1}{4}''$  white, plastic cards for paired-associate presentation.

The apparatus used for the reading task was a Hunter Card Master. The Card Master is designed to present automatically the above plastic cards before a  $3'' \times 5\frac{1}{2}''$  aperture. By means of a timer unit and two metal shutters, the aperture opening can be controlled to secure the following exposure intervals: an anticipation interval to expose the stimulus word alone; a joint presentation interval to expose the word-color pair; and a between-item interval when the aperture is completely covered, during which time the next stimulus-response pair is brought into position behind the shutters.

Four orders of the three word-color pairs appeared in the reading task. Two of these orders presented consecutively constituted a two-trial block. A blank card appeared at

the end of each two-trial block. The following presentation intervals were used: anticipation interval—3 seconds; joint presentation interval—3 seconds; and a between-item interval—2 seconds. All subjects continued in the reading task to a learning criterion of 11 of 12 correct anticipations in two successive two-trial blocks, or for a maximum of 20 trials. Subjects reaching a learning criterion before 20 trials were given perfect scores on the remaining trials.

The reading task was administered immediately following pretraining. At the beginning of the reading task, the subject was shown each color patch with the word concealed and told the appropriate color name. The three color patches were then presented a second time and the subject asked to name the colors. If the subject could not give the correct color names he was eliminated. Following the color naming, a familiarization trial with the three word-color pairs was given. The subject was told with each paired presentation that the word and color went together. After presenting the third pair the subject was instructed that the game was to guess the color that went with each word before the color appeared. The learning trials began following these instructions for subjects in Groups RR, IR, and for half the LD group. Since the LD group would not see the relevant letters as parts of words during pretraining, subjects in this group, presumably, would need to learn where to look for these letters in the reading task words, as well as learn to discriminate the relevant letter from the rest of the word. To control for learning where to look, the other half of the LD group, designated LD-O, was given special orienting instructions at the beginning of the reading task. At the end of the familiarization trial, the experimenter told subjects in the LD-O group always to look at the beginning of the word and this would help them learn the right color. At the same time the experimenter pointed to the beginning of the word. These instructions were repeated twice more, at the end of trial Blocks 1-2 and 3-4.

To secure attention to the stimulus words as they appeared in the aperture during the reading task, the experimenter frequently pointed to the window with a pencil on each presentation. If the subject did not respond during the anticipation interval, the experimenter encouraged the subject to try to guess a color after looking carefully at each word.

If a correct response occurred in the an-

ticipation interval, the subject moved a bead across a counting frame. If an incorrect response or omission occurred, the experimenter named the correct color in the joint presentation interval. The bead incentive and occasional verbal encouragement from the experimenter served as rewards in the reading task.

### Subjects

The subjects were 60 children from three kindergartens in the public schools in Iowa City. The subjects were randomly assigned to the three pretraining groups: 15 each to Groups RR and IR, 30 to Group LD. Three additional subjects were eliminated from the study because they could not respond correctly in naming the color patches. Subjects ranged in age from 61 to 72 months. The mean age was 66.7 months. There were no significant age differences among the three pretraining groups, either in variability, or mean age. Testing was done in the

public schools during November and December.

### RESULTS

#### Pretraining

Comparing Groups RR and IR, the mean number of correct responses were 19.40 and 18.60, respectively; the standard deviations 4.22 and 3.52, respectively. A Lindquist (1956) Type I analysis of variance of the pretraining scores showed no reliable performance differences between these groups. The trials effect for the combined groups was significant ( $p < .01$ ), indicating a learning effect. Since there were no performance differences between Groups RR and IR in pretraining these two word discrimination groups were combined and designated as Group WD. Comparing Groups WD and LD in pretraining, the mean number of correct responses were 19.00 and 21.70, respectively; the standard deviations 3.91 and 1.98, respectively. Hartley's test (Walker & Lev, 1953, pp. 193-194) for homogeneity of variance between the groups was significant ( $p < .01$ ), leading to the rejection of the null hypothesis. The mean numbers of correct responses for Groups WD and LD during pretraining are presented in Table 2. A summary of a Lindquist (1956) Type I analysis of variance for these data is presented in Table 3. As shown in the summary, differences in group means and the trials effect were significant ( $p = .025$ , to adjust for effects of heterogeneity of variance). A further analysis of simple effects showed that the trials or learning effect was entirely accounted for within the WD group ( $p < .01$ ).

#### Reading Task

Comparing Groups RR and IR, the mean number of correct responses were 33.07 and 32.93, respectively; the

TABLE 2  
MEAN NUMBER OF CORRECT RESPONSES  
DURING VISUAL DISCRIMINATION  
PRETRAINING

Group	Blocks of two trials				
	1-2	3-4	5-6	7-8	Total
WD <sup>a</sup>	4.27	4.60	5.03	5.10	19.00
LD	5.47	5.23	5.43	5.57	21.70

<sup>a</sup> Groups RR and IR combined.

TABLE 3  
ANALYSIS OF VARIANCE OF CORRECT  
RESPONSES IN EIGHT LEARNING  
TRIALS FOR GROUPS WD AND  
LD IN VISUAL DISCRIMI-  
NATION PRETRAINING  
(Two-trial blocks)

Source	df	MS	F	p
Between-Subjects	59			
Groups (B)	1	27.34	11.02	<.005
Error (b)	58	2.48		
Within-Subjects	180			
Trials (A)	3	3.19	4.25	<.01
A × B	3	1.98	2.64	<.10
Error (w)	174	.75		
Total	239			

TABLE 4  
MEAN NUMBER OF CORRECT RESPONSES  
AND STANDARD DEVIATIONS FOR  
GROUPS ON THE READING TASK

Group	Mean	SD
WD <sup>a</sup>	33.00	11.67
LD	29.53	12.75
LD-O	32.53	14.59

<sup>a</sup>Groups RR and IR combined.

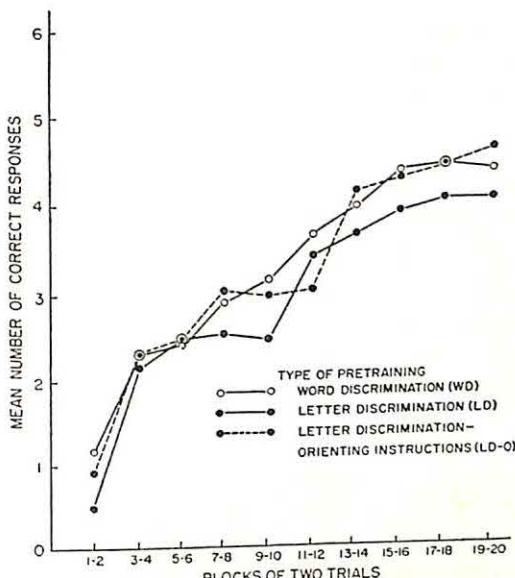


FIG. 1. The effects of visual discrimination pretraining with relevant letters contained in words (Group WD: Groups RR and IR combined) vs. relevant letters presented alone (Groups LD and LD-O) on correct responses in learning to read a three word list.

standard deviations, 11.58 and 11.75, respectively. A Lindquist (1956) Type I analysis of variance showed Groups RR and IR did not differ significantly in overall performance on the reading task. A further comparison between the groups on the first trial block also showed no reliable differences.

On the basis of these findings, the two word-discrimination pretraining groups were again combined and designated as Group WD in the reading task. The mean number of correct responses for Groups WD, LD, and LD-O are presented in Table 4. The

learning curves in the reading task for the three groups are presented in Figure 1. A summary of a Lindquist (1956) Type I analysis of variance for these data showed no significant overall performance difference for the groups on the reading task. Comparing Group LD with Groups WD and LD-O separately showed no reliable performance differences ( $F < 1$ ). Per-

formance differences on the first trial block also were not significant.

#### DISCUSSION

If word shape is the relevant visual aspect for discriminating among words for beginning readers, Group RR pretraining should have resulted in transfer of discrimination learning to the reading task and faster learning of the list relative to Group IR pretraining. On the other hand, if the relevant letters alone provide the basis for visual discrimination in pretraining, Groups RR and IR pretraining should

have resulted in equal transfer of discrimination learning and no performance difference between the groups on the reading task.

The findings based on comparing Groups RR and IR on the reading task are consistent with the assumption that beginning readers discriminate among similar length words of different shape on the basis of specific letter differences. Further investigation is necessary to determine the effects of at least two other word-stimulus variables on visual discrimination learning: word-shape differences defined in terms of differences in word length and location of the relevant letter within the stimulus word, i.e., in other positions than at the beginning of the word. It is possible that placing the relevant letter in some middle position in the word would increase the difficulty of the letter discrimination. This effect could increase the importance of shape or length cues as bases for discrimination.

The findings based on comparing Groups WD (Groups RR and IR combined), LD, and LD-O on reading task performance require an explanation to account for the appearance of learning differences between the word and letter discrimination groups in pretraining, and the absence of learning differences on the reading task. Learning differences in discrimination pretraining indicated differential difficulty in discriminating letters when these letters were presented singly as compared to being presented in words. Absence of learning differences on the reading task indicated that the addition of irrelevant letters to the previously differentiated single-letter stimuli for Groups LD and LD-O did not increase the difficulty of discriminating these letters.

To account for the relative difficulty of the WD group discrimination learn-

ing in pretraining, it is assumed that generalization among the nonsense words increased as a result of the addition of similar stimulus elements—the same two letters—to the relevant letters in each of the words. Thus to discriminate among the words, WD group subjects had to acquire a set of observing responses related either to the location of the relevant letter in the words, discriminating the relevant letter from the total stimulus complex, or a combination of the two types of observing responses.

On the basis of this reasoning, however, differences in learning performance should have appeared in the reading task, at least for the LD group. The assumption would be that the LD group would need to acquire the same orienting responses to the relevant letters included in the reading task words. This additional learning would be predicted to interfere with the acquisition of the word names. The findings did not confirm this expectation.

To account for the absence of a WD-LD group performance difference on the reading task, two explanations are offered. The simplest explanation is based on the greater difficulty and variability of the reading task. The increase in performance variability could mask any differential transfer of discrimination learning between the pretraining groups. A second explanation requires the assumption that exposure to the relevant letters in the LD group pretraining resulted in an "acquired distinctiveness" of the letter stimuli, not detected by the pre-training response measure, that transferred to the reading task. Transfer of this acquired distinctiveness could facilitate the acquisition of the observing responses necessary to discriminate the relevant letters included in the reading task words—the same

observing responses acquired more slowly by the WD group in pretraining.

The second explanation suggests the hypothesis that although letters included in words are more difficult to discriminate than the same letters presented singly, this relative difficulty in learning to discriminate letters in words can be overcome by providing visual discrimination training with the relevant letters prior to presenting them as parts of words.

Further research is necessary to test this hypothesis when the relevant letters are included in other than the first position in the word.

### SUMMARY

The purpose of this study was to determine the stimuli relevant to visual discrimination among word forms for beginning readers.

Of three groups of kindergarten children, two received discrimination pretraining with words. For one group, the pretraining words were constructed so that both shape and letter differences among the words were relevant to discriminating among the words in the reading task. For the second group, only the letter differences among the pretraining words were relevant to discriminating among the reading task words. A third group received discrimination pretraining with the relevant letters alone.

There were no significant performance differences between the two word discrimination groups on either a pretraining or reading task, a finding consistent with the assumption that children discriminate among similar length words of different shape on the basis of specific letter differences.

When the letter discrimination pretraining group was compared with the

combined word discrimination pre-training groups, there was a performance difference favoring the letter discrimination group in pretraining, but no differences on the reading task between the groups. These findings suggested the hypothesis that although letters included in words are more difficult to discriminate than the same letters presented singly, this relative difficulty in learning to discriminate letters in words can be overcome by providing visual discrimination training with the relevant letters prior to presenting them as parts of words.

### REFERENCES

- BUSWELL, G. T. Fundamental reading habits: A study of their development. *Suppl. educ. Monogr.*, 1922, No. 21.
- DAVIDSON, HELEN P. An experimental study of bright, average, and dull children at the four-year mental level. *Genet. psychol. Monogr.*, 1931, **9**, 119-289.
- GATES, A. I., & BOEKER, ELOISE. A study of initial states in reading by preschool children. *Teach. Coll. Rec.*, 1923, **24**, 469-488.
- GIBSON, ELEANOR J., & OLUM, VIVIAN. Experimental methods of studying perception in children. In P. H. Mussen (Ed.), *Handbook of research methods in child psychology*. New York: Wiley, 1960. Pp. 311-373.
- GRAY, W. S. *The teaching of reading and writing: An international survey*. (Monographs on Fundamental Education, No. 10) New York: UNESCO, 1956.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
- MEEK, LOIS H. A study of learning and retention in young children. *Teach. Coll. Contrib. Educ.*, 1925, No. 164.
- MUEHL, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. *J. educ. Psychol.*, 1960, **51**, 217-221.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received February 9, 1961)

# *Psychology and Mental Health*

A report of the Institute on Education and Training  
for Psychological Contributions to Mental Health,  
held at Stanford University in August, 1955.

*Edited by CHARLES R. STROTHER*

**Topics discussed at the Institute:**

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: \$1.75

---

*Order from:*

American Psychological Association, Inc.  
1333 Sixteenth Street, N. W.  
Washington 6, D. C.

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

October 1961

Number 5

## PERSONALITY FACTORS AMONG SCIENCE AND TECHNOLOGY FRESHMEN<sup>1</sup>

EUGENE B. NADLER      AND  
*Case Institute of Technology*

GILBERT K. KRULEE  
*Northwestern University*

In spite of numerous criticisms of the concept of personality types, ably summarized by Hilgard (1957, pp. 485-488), the notion has proved useful in psychology and anthropology under the names of syndrome, life style, basic personality structure, modal personality, etc. (See Inkeles & Levinson, 1954, for example). The study reported here had as its starting point such a conception, one which could be used to differentiate among freshman students in a school of science and technology.

The focus was upon two dichotomous personality variables distinguished by Murray (1938, pp. 222-223), Extratection-Intraception and Exocathexion-Endocathexion. The first refers to the direction of the person's interest, outward or inward, while the second refers to its objects, things, or ideas. Taking all combina-

tions of both the direction and the object of interest led Murray to describe the following four types:

1. Extratection-Exocathexion: To adapt to the world as it stands; to be interested in tangible results; to be very practical; to amass a fortune. To secure a permanent position; to become a member of clubs and institutions; to be without illusions; to conserve established values. To work effectively with mechanical appliances.

2. Intratection-Exocathexion: To live imaginatively; to dramatize the self; to express one's sentiments and beliefs in action. To initiate and further progressive social movements. To speak against abuses; to propose reforms. To concoct new schemes, business ventures, political innovations; to be guided by a vision of the future. To seek adventure; to become involved in amorous affairs.

3. Extratection-Endocathexion: To be interested in ideas and theories about substantial events (for example, physical sciences). To reflect and write about external occurrences and systems; history, economics, government, education. To collect data and think inductively.

4. Intratection-Endocathexion: To devote oneself to artistic or religious representations. To dream, brood, and introspect; to become absorbed in the attempt to resolve inner conflicts and spiritual dilemmas.

<sup>1</sup>This research was part of a larger series of studies on curriculum evaluation supported by a grant from the Carnegie Corporation of New York. Additional support was made available through a faculty research fellowship awarded to G. K. Krulee by the Social Science Research Council. The authors gratefully acknowledge the support and encouragement of the Case Institute of Technology administration. We are also very much in debt to Barbara Doty who assisted in the collection and analysis of the data and to Morrell Heald for his cooperation in the administration of the questionnaires. The opinions expressed herein are the authors'.

The study was concerned with the following tasks:

1. testing the hypothesis that significant numbers of students in the school would fall into the two extraceptive categories, and very few into the two intraceptive categories

2. placing the predicted finding in a broader societal context by determining what socioeconomic influences, if any, might be reflected in the two varieties of extraception

3. describing further the differences between the extraceptive students who were exocathecting on the one hand and endocathecting on the other in terms of their career goals and life values

4. examining the influence of these orientations on the students' evaluation of their courses and curricula, and on their academic performance

### METHOD

*Differentiating the types.* The following questions were included as part of a larger questionnaire<sup>2</sup> presented to 432 male freshmen about three-quarters of the way through their first semester:

1. Once in my life I would like to: a. develop a valid scientific theory, b. discover a satisfactory principle of ethical conduct, c. invent a marketable product, d. convince a legislature to enact one of my views into law.

2. I would like a job where: a. I could express my social views through my work, b. I could organize the thinking in my field in a systematic way, c. I could discover my true potentialities, d. I could do work of practical importance.

3. I am filled with awe when I consider the lives of: a. Vincent Van Gogh, Mahatma Gandhi, Albert Schweitzer; b.

Charles Kettering, Andrew Carnegie, Winston Churchill; c. Louis Pasteur, Woodrow Wilson, Abraham Lincoln; d. Frank Lloyd Wright, John Dewey, Albert Einstein.

4. In my spare time I would like to: a. teach woodcraft to a Boy Scout troop, b. participate in a campaign for better schools, c. discuss and evaluate different political systems, d. learn about psychoanalysis.

5. Everyone is a little of everything, but by and large, I am mostly: a. practical, b. idealistic, c. thoughtful, d. sensitive.

For every question, each of the four alternatives corresponds to one of the four personality types. If in answering the five questions a student chose three or more alternatives characteristic of a single type, he was placed in that type. The chance probability of this happening was  $p = .10$ , calculated by collecting the first three terms of the binomial expansion ( $\frac{1}{4} + \frac{3}{4})^5$ .

*Questionnaire.* The same questionnaire included a large number of items which covered the general areas of social class background, career goals and life values, and evaluation of curricula and courses.

*Academic performance.* At the end of the first freshman semester, two measures of achievement became available, grade point averages and the number of separations.

### RESULTS

Of the 432 students surveyed, 143 or 35% were classed as Extranceptive-Exocathecting, and 83 or 19% were classed as Extranceptive-Endocathecting. Only 15 men fell into the Intranceptive-Exocathecting category (3.4%), and only 9 into the Intranceptive-Endocathecting category (2.1%). The miscellaneous or mixed group (those with fewer than three responses in their largest category) consisted of 182 individuals, comprising 42% of the sample. Thus, the two Extranceptive groups who were predicted to make up a significant proportion of the sample together accounted for about 52% of the subjects, with only insignificant numbers in the Intranceptive categories. The rest of this report will be solely concerned with the two sub-

<sup>2</sup> A copy of this questionnaire has been deposited with the American Documentation Institute. Order Document No. 6834 from ADI Auxiliary Publication Service, Library of Congress; Washington 25, D. C., remitting in advance \$2.00 for microfilm or \$3.75 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

groups among the Extraceous men, the Exocatheters and the Endocatheters, who will be referred to as the practical men and the theoretical men, respectively, for convenience and descriptive simplicity.

### Social Class Background

The responses of the subjects to a question dealing with their fathers' occupations at the time of their birth (about 1940) and to one dealing with their fathers' occupations at the time of taking the questionnaire (early 1959) were sorted into the following categories: (a) unskilled, semiskilled, farm; (b) skilled; (c) foremen; (d) white collar, sales, government, military; (e) subprofessional; (f) entrepreneurial; (g) professional. Table 1 shows the percentage distributions of each group of fathers at the two times.

The chi square test of independence between the two distributions, practical versus theoretical fathers in 1940, was 15.97, significant at the .02 level. The differences between the two distributions were due to the higher proportion of unskilled, semiskilled, farm workers, and foremen among the fathers of the practical men, and the relatively higher proportion of white collar, sales, government, and military workers among the fathers of the theoretical men.

The distributions of fathers' occupations were still significantly different in 1959, chi square being 84.50, significant beyond the .001 level. However, the major difference here was due to the larger proportion of entrepreneurs among the fathers of the theoretical men.

Comparing each group with itself at the two different times showed a major shift among the practical fathers from unskilled, semiskilled, and farm in 1940 to white collar, sales, government, and military occupations

TABLE 1  
FATHER'S OCCUPATION AT DIFFERENT TIMES

Father's occupation	Practical men (%)		Theoretical men (%)	
	1940	1959	1940	1959
Unskilled, semiskilled, farm	36	18	20	11
Skilled	11	13	16	19
Foremen	8	5	0	1
White collar, sales, government, military	18	35	29	30
Subprofessional	3	1	7	0
Entrepreneurial	8	12	12	26
Professional	16	18	17	14
	x <sup>2</sup>		p	
Chi square comparisons:				
Practical vs. theoretical: 1940	15.97		.02	
Practical vs. theoretical: 1959	84.50		.001	
Practical: 1940 vs. 1959	18.56		.01	
Theoretical: 1940 vs. 1959	11.16		.07	

in 1959; the chi square between the 1940 and 1959 distributions was 18.56, significant beyond the .01 level. The fathers of the theoretical men, on the other hand, showed a major increase in the entrepreneurial category; the chi square of 11.16 was not significant, but was close enough to significance to indicate a slight trend, p being about .07. It is interesting that the occupational distribution of the fathers of the practical men in 1959 is almost identical with that of the fathers of the theoretical men in 1940.

There were no significant differences in fathers' incomes between the two groups, and the omission of almost 40% of the responses on questions dealing with parental education invalidated any conclusions that might have been drawn from these data.

### Career Goals and Life Values

In order to analyze the curriculum choices of the students, the school curricula were divided into science and engineering. The one curriculum which did not fit into either category, management, was safely excluded since only one practical man and one theoretical man chose it. The science curriculum comprised math, physics, chemistry, and engineering science. The engineering category included metallurgical, civil, chemical, electrical, and mechanical engineering. In this way a two by two table was con-

structed with personality type, practical or theoretical, on one margin, and curriculum choice, engineering or science, on the other. The chi square was 59.24, significant beyond the .001 level. Eighty-four percent of the practical men had chosen engineering, 16% science, as opposed to 67% of the theoretical men who had chosen science and 33% engineering.

The distribution of choices of the ideal job upon graduation for the practical men was significantly different from that of the theoretical men, chi square being 41.25, significant beyond the .001 level. The order of choices for the practical men was graduate school, management, equipment design, applied research, new product development, and bringing up the distant rear, basic research, production, sales, and teaching. The order of choices for the theoretical men was graduate school first, and all other alternatives far behind.

The requirements of the ideal job also differentiated the two groups. Out of a list of 10 possible requirements of the ideal job which were checked for "high" priority, four items discriminated between the two groups, the probabilities ranging between .05 and .001. Among the practical men it was more important to earn a good deal of money, whereas among the theoretical men it was more important to have the opportunity to use their special abilities and aptitudes, to be creative and original, and to be relatively free of supervision by others. Out of the same list which was rechecked for items having "medium" priority, five items separated the groups, the probabilities ranging between .05 and .02. Among the practical men it was moderately important to have the opportunity to use their special abilities and talents, to be creative and original, and to exercise leadership, whereas it

TABLE 2  
CHARACTERISTICS OF JOB AND LIFE GOALS  
CHOSEN AS IMPORTANT

Characteristic of life goal	Practical men	Theoretical men
Ideal job:		
Management	17%	4%
Equipment design	14	4
Graduate school	30	73
Characteristics of ideal job:		
Use of special abilities	62	81
High income	50	28
Opportunity for creativity	51	76
Freedom from supervision	30	44
Ideal position at peak of career:		
Science	25	41
Administration	50	27
Teaching	1	9
Personal characteristics thought important:		
To be liked by people	47	27
To understand other people	56	35
Good grounding in basic theory	59	73
Practical knowledge of facts	78	61
Ability to convince and persuade	40	24
Organizing and administrative ability	47	30
Activities expected to give the most satisfaction:		
Career	22	36
Family	66	41
Community and organizations	2	10

was moderately important for the theoretical men to have a chance to earn a good deal of money and to look forward to a stable, secure future.

Responses to a question about the occupational positions which would be ideal at the peak of their careers were classified into categories of scientific, administrative, ownership, teaching, and other and don't know. The chi square test of independence between the two distributions was 78.17, significant beyond the .001 level. Among the practical men the order of choices was administrative, scientific, ownership, and teaching. Among the theoretical men this order was scientific, administrative, teaching, and ownership.

The personal characteristics which the two groups perceived as promoting vocational success also showed some differences. Of 17 items checked as essential to vocational success, six distinguished the two groups at probability levels ranging from .05 to .01. The practical men felt it important to be liked by people, to understand other people, to be able to convince and persuade other people, to have practical knowledge of the facts in one's field, and to have organizing and administrative ability. The theoretical men more often felt it necessary to have a good grounding in basic theory.

A short series of four questions dealing with the importance of planning for the future, getting ahead in life, knowledge of one's academic performance in comparison to others, and choosing among being successful, independent, or well liked showed one significant difference between the two groups, significant at the .01 level. The practical men preferred success to independence, and independence to being well liked, whereas the theoretical men preferred independence to success, and success to being well liked.

The activities which were expected to give the most satisfaction in life showed the practical men overwhelmingly making family relationships a first choice, followed by career activities and leisure time activities. The theoretical men were about evenly divided between family relationships and career activities, with some sentiment for activities directed toward national and international betterment, and church activities. The chi square between the two distributions was 17.74, significant beyond the .01 level.

#### *Evaluation of Curricula and Courses*

In 5 out of 13 ratings describing the nature of the ideal curriculum, the two groups presented different conceptions; four of these differences were significant beyond the .001 level and one at the .01 level. The practical men thought the ideal curriculum should emphasize practical aspects of the material, relate directly to the kind of work they would be doing, emphasize facts and techniques that have direct application, leave enough free time for other activities, and prepare them for taking responsibilities in management and administration. The theoretical men thought that the ideal curriculum should be mostly concerned with theory, provide them with a broad and general knowledge of the field, emphasize the ability to think logically and abstractly, be difficult enough to require real effort, and not have to prepare them for taking responsibilities in management and administration.

Two out of 13 ratings describing what the students realistically expected of their curricula distinguished the two groups, one being significant at the .001 level and one at the .01 level. The theoretical men expected their chosen curriculum to be mostly concerned with theory, but did not

TABLE 3  
MEAN RATINGS OF ITEMS RELATING TO  
THE CURRICULUM

Item <sup>a</sup>	Practical men	Theoretical men
The ideal curriculum should:		
1. be mostly concerned with theory (1) <sup>b</sup>	5.72	3.74
emphasize practical aspects (8)		
2. provide broad knowledge (1)	4.15	3.15
relate directly to future work (8)		
5. emphasize the ability to think (1)	4.42	2.69
emphasize facts and techniques (8)		
6. require real effort on my part (1)	3.78	2.57
leave me with free time (8)		
10. prepare me for management (1)	2.92	4.16
not prepare me for management (8)		
My realistically expected curriculum would:		
1. be mostly concerned with theory (1)	4.80	3.32
emphasize practical applications (8)		
10. prepare me for management (1)	3.70	4.70
not prepare me for management (8)		
In my least valuable course:		
1. practical applications are emphasized (1)	5.10	3.62
the primary emphasis is on theory (8)		
3. students influence what is discussed (1)	5.57	6.86
instructor decides what is discussed (8)		
11. the instructor is impartial (1)	3.13	2.36
the instructor is not impartial (8)		

<sup>a</sup> All items are paraphrased.

<sup>b</sup> Numbers in parentheses are the extreme scale values.

expect it to prepare them for taking on responsibilities in management and administration; the opposite was true of the practical men.

Neither the nominations for the most valuable course, nor the descriptive ratings of the most valuable course distinguished the practical men from the theoretical men.

On the other hand, the nominations for least valuable course were different for the two groups, the chi square of 12.926 having a probability of less than .05. Whereas 44% of the theoretical men nominated graphics as the least valuable course, only 17% of the practical men did. On the other hand, 30% of the practical men chose chemistry as their least valuable course, as opposed to 24% of the theoretical men. Other courses were chosen with equal relative frequency by both groups. In 3 out of 14 ratings describing the least valuable course, the two groups presented different evaluations, each significant beyond the .005 level. The theoretical men described their least valuable course as primarily emphasizing practical aspects of the material and were more vehement in expressing the idea that the instructor did not allow the students to decide what to discuss in class. However, they also felt he was more impartial in dealing with individual students than did the practical men.

#### Academic Performance

The grade point average of the practical men at the end of the first semester was 1.034 on a 3-point grading system which included 2 negative quality points for F. The average of the theoretical men for the same period was 1.378. The difference between these two means was significant beyond the .01 level for a two-tailed test,  $t$  being 2.991.

Since it was possible that the obtained difference in school achievement was due to differences in scholastic aptitude between the two groups, they were compared with respect to their performance on two tests of the

College Entrance Examination Board required of all freshmen. There was no significant difference between the two groups on the Math Aptitude test. On the Verbal Aptitude test, the theoretical men scored slightly higher than the practical men;  $t$  was 1.83, significant at the .05 level for a one-tailed test but not for a two-tailed test. Thus, there was only weak support at best for the notion that the higher achievement of the theoretical men was due to higher college aptitude.

There was no significant difference in separations between the two groups at the end of the first semester, the percentage being 22% for the practical men and 20% for the theoretical men. However, on the hypothesis that personality compatibility or ego syntonicity (see Levinson, 1959) might play a role in the separation rate, the subjects were divided between those who had made compatible choices and those who had not. Practical men who chose engineering curricula, or theoretical men who chose science curricula were considered to have made compatible choices; practical men who chose science, and theoretical men who chose engineering were considered to have made incompatible choices. Of 174 men who had made compatible choices, 30 or 17% were separated at the end of the first freshman semester. Of 50 men who had made incompatible choices, 17 or 34% were separated at the end of the first freshman semester. The difference between these proportions was significant beyond the .01 level.

#### DISCUSSION

The major differences between these Extrareceptive freshmen may perhaps be best understood in terms of the social psychological concepts of *status anxiety* and *status security*. The pattern of responses to the questionnaire, and especially the data on father's oc-

cupation, suggests that the practical men were relatively more anxious about their status, while the theoretical men were relatively more status secure. The orientations of the practical men tended to serve the ends of upward status mobility in modern industrial society; those of the theoretical men gave greater personal and horizontal mobility. The courses and curricula of the school were evaluated in terms of whether they met the requirements of these orientations.

In the days when the differences between science and engineering were sharp and clear these orientations might have had approximately equal viability. Boys who liked to work with their hands could go into engineering; thoughtful boys could go into science.

The trend in recent years, however, has been for the engineering curricula to become more like the science curricula, to emphasize basic scientific principles and broad theoretical skills. The first year curriculum facing the students in this study had this marked theoretical character, and was identical for all of them. As such, it was less ego syntonic for the practical men than for the theoretical men, which probably accounts for the lower grade point average of the practical men at the end of the first semester.

In the first year, all the practical men faced a program that was not ego syntonic. However, for those who had chosen an engineering curriculum, the first year was at least a stepping stone to a program that was more compatible with their needs. No such hope remained for those practical men or those theoretical men who had made incompatible curriculum choices, since the feedback from upperclassmen and faculty members regarding the nature of the second, third, and fourth year programs could only increase the conflicts arising from their incongruous decisions. Thus, at the end of the first

freshman semester, those who had made incompatible curriculum choices had double the separation rate of those who had not.

What practical recommendations may be made from these considerations?

In general, the specific recommendations tend to fall under two main headings: (a) techniques designed to help students make ego syntonic choices at the outset, and (b) techniques designed to make it easier to change from an incompatible program to one that is compatible.

The present state of personality measurement probably does not permit that the research tool used here to classify freshman subjects could also be used to select prospective freshmen, even if it were greatly improved. On the other hand, such an instrument would have great value in aiding the preliminary counseling of freshmen. Although the subjects of this study were not required to declare a choice of curriculum until the end of the first year, and had an opportunity to revise this choice even afterward, it is clear that sometime during the first semester most of the students had made a fairly binding decision, since almost all of them answered the question on curriculum choice, whereas other questions went unanswered. A guidance instrument would prove of value in clarifying for the student the consequences of given curriculum decisions, and could aid in reversing seemingly binding decisions which were incompatible with the student's interest.

As a further way of improving the initial choices of freshmen, the publicizing departments of technical schools could step up a campaign to present a more realistic picture of what a modern technical program entails, focusing on the emphasis on theoret-

ical skills. The present study indicates that, by and large, there is a great deal of self-selection in the students who come to a technical school, and that, by and large, this self-selection is highly realistic. An informative publicizing campaign could improve an already effective self-selection system. It also might have the effect of attracting students who might otherwise invest their energies elsewhere, thus influencing another side of the selection process.

There remains the problem of changing incompatible choices once they have been made. Our impression is that the student's decisions have been accepted by members of his family and his friends to a high degree, so that part of his identity has come to rest on the expectations that he has induced in others. It would appear to be desirable for the school to counter this social support, to create a climate in which indecision, hesitancy, and ambivalence about curriculum is tolerated, in which shopping around is encouraged, and in which curriculum choices may be more easily postponed or changed.

A second recommendation which attacks this same problem is one which is already being carried out in some respects. With the breakdown of the traditional boundaries between science and engineering in the working world, there is some tendency for a greater number of course choices to be offered the student, thus increasing the possibility of accommodating both practical and theoretical interest patterns within the same department.

#### SUMMARY

Measuring Intraception-Extraception and Exocathection-Endocathection in the freshman class at a school of science and engineering showed almost no Intraceptive persons present.

The Exocathecting men tended to choose engineering as a major field of study, were oriented toward the application of social skills in managerial and administrative occupations as means of achieving money and success, and perceived family relationships as bringing satisfaction in life. The Endocathecting men tended to choose science as a major field of study, were oriented toward utilizing theoretical skills in graduate school and scientific occupations as means of achieving self-actualization, and perceived family and career activities as jointly bringing satisfaction in life. In general, the Exocathecting men seemed more anxious about their social status than the Endocathecting men. Compatibility between personal needs and objective features of the school situation was seen as a major factor in the higher grade point average of the En-

docathecting men, and in the lower separation rate of those students who had made ego syntonic curriculum choices. Some practical recommendations stemming from these findings were briefly discussed.

#### REFERENCES

- HILGARD, E. R. *Introduction to psychology*. (2nd ed.) New York: Harcourt, Brace, 1957.
- INKELES, A., & LEVINSON, D. J. National character: The study of modal personality and sociocultural systems. In G. Lindsey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- LEVINSON, D. J. Role, personality, and social structure in the organizational setting. *J. abnorm. soc. Psychol.*, 1959, **58**, 171-180.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1958.

(Received September 13, 1960)

## THE DRAW-A-MAN TEST IN INDONESIA<sup>1</sup>

R. MURRAY THOMAS

*University of California, Santa Barbara*

AND

ANWAR SJAH

*Padjadjaran University, Bandung*

Indonesia, a nation of perhaps 90 million people, has no standardized intelligence scale. This state of affairs is the result of a variety of factors, including a dearth of trained test constructors, the existence of many different languages in the nation, and the impossibility of establishing accurate birthdates for many subjects who might form a standardization group.

In trying to provide a temporary instrument to use with children until an adequate test is developed, the writers tried the Goodenough Draw-A-Man scale on a sample of 1,245 children in Bandung, a modern city of about one million inhabitants in the mountains of West Java.

The results of this study are reported here as: (a) a comparison between the success of the Indonesian children and the success of the American children in Goodenough's original standardization group to discover whether American standards for the test should be transferred directly for use in Indonesia, (b) correlations between the Indonesian children's Draw-A-Man scores and their teachers' ratings of their general intelligence, and (c) correlations between the Indonesian pupils' Draw-A-Man scores and their marks in several school subjects to discover whether the test correlated with achievement in school subjects.

<sup>1</sup>This study was carried out under a teacher education project of the State University of New York financed by the Ford Foundation. The aid of Winarno Surachmad and research assistants Soeparman, Mariani Azis, Suharsini, and S. Hutabarat is gratefully acknowledged.

### METHOD

#### Subjects

The 1,245 children were drawn from three sources: 639 children from five public elementary schools enrolling Grades 1-6, 524 children from five private schools enrolling Grades 1-6, and 82 children from three private kindergartens (there are no public kindergartens).

The schools were selected by random sampling techniques from the Bandung district, and every child in class the days of the testing (late spring 1960) was included. In the final compilation of data 109 children were eliminated from the sample because their birthdates could not be established, and 28 were eliminated because they drew women or included cars or boats instead of drawing only a man. This reduced an original sample of 1,382 to a usable sample of 1,245.

The nation's ideal of compulsory, universal education has not been realized, primarily because of a shortage of school buildings. Thus not all Bandung children attend school. An inspection of the occupations of the fathers of school groups like this sample suggests that it is biased somewhat in favor of the middle and upper socio-economic classes.

#### Procedure

In each classroom the research assistant who administered all of the tests passed letterhead size paper to the subjects and asked them to draw, with pencils, a picture of a man. While the pupils drew, their teacher was asked to make a judgment of each

child's overall intelligence and to place him in one of seven graduated categories of intelligence. (The Indonesian term *ketjerdasan* was used to mean *intelligence*.) The investigators' purpose was to learn whether there was a relation between scores on the Draw-A-Man Test and intelligence, according to the teacher's own estimate of each pupil's overall ability. The seven intelligence categories were named: very highest, quite high, somewhat above average, average, somewhat below average, quite low, and very lowest. The teachers were instructed to think of these levels as referring to school children; so the lowest category was for the least intelligent children found in school, not the least intelligent in the general population.

### RESULTS

Table 1 compares the Indonesian pupils with the normal-school-progress group from Goodenough's original standardization sample. (Goodenough compared an advanced-progress group with normal-progress and retarded-progress groups but she based the test's standards only on the normal-school-progress group.)

Although the Indonesian data were at first separated for analysis into private and public school samples, the means and standard deviations for the private and public schools were found to be almost identical at each age level. Since it appeared that the two groups represented the same population, they were combined for comparison with the American sample.

As shown by Table 1, the Indonesian group's mean was above the Goodenough-normal-progress mean at ages 5 (difference significant at .01 level) and 6 (significant at .05 level). But beyond age 6 the Indonesian means lagged behind the Goodenough-nor-

mal-progress means at all ages from 7 through 12 (all significant at .01 level) with the exception of age 11 where the Indonesian group spurred ahead of the Goodenough group a slight but statistically insignificant amount. (At no age did the Indonesian average rise as high as the Goodenough - advanced - school - progress group or fall as low as the Goodenough - retarded - school - progress group.)

Indonesian scores on the Draw-A-Man Test at each age level showed significantly more variation than did the Goodenough - normal - progress scores. These differences in variances were significant at the .01 level for ages 6, 7, 8, 10, and 11. They were significant at the .05 level for ages 5, 9, and 12 (Table 1). Perhaps this is due to the fact that by definition the Goodenough-normal-progress group did not contain the extremes that would be found in an unstratified sampling of school children.

In Table 2 correlations (*r*) between Draw-A-Man scores and teachers' ratings of intelligence are reported for each grade level, kindergarten through sixth. The correlations were all positive and moderately low or, as in Grades 2 and 4, quite low.

Table 3 shows correlations (*r*) between pupils' Draw-A-Man Test scores and the most recent end-of-the-semester marks received by the pupils in various school subjects. (The Indonesian marking system uses numbers ranging from 1 to 10, with teachers also commonly adding marks halfway between each pair of numbers, such as 5½. The extreme marks of 1, 2, 9, and 10 almost never appear.)

All correlations were positive. They ranged from very low (.04) to rather high (.80). In general the highest correlations were between the Goode-

TABLE 1  
COMPARISONS OF INDONESIAN WITH AMERICAN GROUP

Group	Age at last birthday								
	5	6	7	8	9	10	11	12	13
Indonesian:									
<i>N</i> cases	35	104	135	202	204	183	165	121	98
<i>M</i>	13.6	15.1	16.8	20.0	23.3	26.9	33.2	32.2	36.9
Median	12.4	14.7	16.3	19.5	22.4	24.9	34.8	32.7	37.5
<i>SD</i>	5.3	5.5	6.0	6.3	7.9	8.3	8.5	8.6	6.8
Goodenough-normal-progress:									
<i>N</i> cases	623	802	669	503	504	351	101	22	
<i>M</i>	10.2	13.9	18.5	22.5	26.2	30.6	33.0	35.2	
Median	9.9	13.9	17.8	22.3	25.8	30.7	33.1	36.1	
<i>SD</i>	4.1	4.4	4.7	5.4	7.0	6.6	6.6	6.2	

nough test and arithmetic marks. Next highest were those in physical science (not taught below Grade 5), geography, and biology. On the average,

correlations with reading, Indonesian language (the national language), and Sundanese (the local dialect) were somewhat lower.

TABLE 2  
CORRELATIONS OF GOODENOUGH TEST SCORES WITH TEACHER RATINGS OF INTELLIGENCE

	Grade level						
	<i>k</i>	1	2	3	4	5	6
<i>N</i>	82	210	238	221	218	151	97
<i>r</i>	.22*	.32**	.07	.27**	.11	.29**	.22*

\*  $p < .05$ .

\*\*  $p < .01$ .

### SUMMARY AND CONCLUSIONS

In Bandung, Indonesia, the 1,245 school children, kindergarten through Grade 6, who were given the Goodenough Draw-A-Man Test did not, on the average, score as high as did the American group on which Goodenough originally based her standards for the test, suggesting that the American standards for mental ages should not be applied to Indonesians.

When Goodenough test scores were

TABLE 3  
CORRELATIONS OF DRAW-A-MAN SCORES AND SUBJECT MATTER MARKS

Subject matter	Grade level					
	<sup>1</sup> ( <i>N</i> = 210)	<sup>2</sup> ( <i>N</i> = 237)	<sup>3</sup> ( <i>N</i> = 217)	<sup>4</sup> ( <i>N</i> = 214)	<sup>5</sup> ( <i>N</i> = 158)	<sup>6</sup> ( <i>N</i> = 105)
Arithmetic	.42**	.52**	.42**	.55**	.56**	.80**
Reading	.43**	.22**	.54**	.29**	.22**	.20*
Indonesian	.07	.31**	.22**	.15*	.24**	.32**
Sundanese	.26**	.06	.25**	.35**	.27**	.04
Geography			.25**	.30**	.44**	.64**
Biological science				.18**	.27**	.55**
Physical science					.44**	.69**

\*  $p < .05$ .

\*\*  $p < .01$ .

correlated with school marks in various subject matter areas, the median correlation was about .30. The highest relationships were found between the Draw-A-Man Test and arithmetic marks (from  $r = .42$  to  $r = .80$ ). Correlations between Draw-A-Man scores and teachers' judgments of children's intelligence were low, ranging from .07 to .32 for different grades.

The writers concluded that the Goodenough test yielded sufficiently

promising results to warrant its use as a measure of ability for Indonesian children (especially for estimating arithmetic and science abilities in upper elementary grades) until a more adequate test has been developed.

#### REFERENCE

- GOODENOUGH, FLORENCE L. *Measurement of intelligence by drawings*. Yonkers, New York: World Book, 1926.

(Received December 21, 1960)

## AMERICAN GRADUATE STUDENTS: SOME CHARACTERISTICS OF ASPIRING TEACHERS AND RESEARCHERS

DAVID GOTTLIEB

*Department of Sociology and Anthropology, Michigan State University*

In recent years, increasing attention has been paid to the phenomena of adult socialization as observed within American colleges and graduate schools (Becker & Strauss, 1956; Gottlieb, 1961; Jacob, 1957; Merton, Reader, & Kendall, 1957). While the ratio of speculation to empirical findings in this field is high, the available material would indicate that changes do occur in the values, attitudes, and behavior patterns of students in training and that these changes may at least in part be attributed to the social system of our educational institutions. Several investigators (Becker & Geer, 1958; Spaeth, 1959) have attempted to note how changes in the occupational values expressed by graduate students may be tied in with a theory of "anticipatory socialization" (Merton & Kitt, 1955), a process by which the individual adopts the values of a group to which he aspires but does not yet belong.

### Purpose

The research being reported here deals with a comparison of two groups of American graduate students: (a) the student who comes into the graduate school system with a preference for a career in either teaching or academic research and reports that there has been *no change* in his preference as he has moved through professional training, (b) the student who states that since he has been in graduate school either teaching or research has become more attractive as a post-graduate career choice. The major

purpose of this study was to investigate how these students, who find themselves at different stages of socialization and career commitment, vary in respect to certain values and attitudes.

### METHOD

*Subjects.* The National Opinion Research Center's sample of graduate students from which the data for this analysis has been taken is a two-stage stratified cluster sample of the enrollment of American graduate schools which offered the PhD in the "traditional arts and sciences" in the autumn of 1958. A total of 2,842 students from 25 graduate schools participated in this survey.<sup>1</sup>

*Procedure.* In order to obtain some measure of the source and direction of career change each student was asked the following:

Please rank the following in terms of your personal preference as a future occupation.

- a. Teaching undergraduates
- b. Teaching graduates
- c. Doing research in your field
- d. Academic administration

Have your opinions on these alternatives changed since you first decided to go to graduate school? Please write the letter or letters of alternatives which seem more desirable to you now.

The responses to this set of questions were used as the basis for a change index. Persons who indicated that any alternative was more desirable to them are considered "changers"; persons who indicate no change in preference are considered "nonchangers." It was found that 41% of the graduate students in this

<sup>1</sup> The research reported here is part of a larger investigation of American graduate students conducted by the National Opinion Research Center. I am indebted to James A. Davis, Study Director, and to Joe L. Spaeth, Assistant Study Director, for their continued cooperation.

TABLE 1  
CAREER PREFERENCES AND CHANGE STATUS<sup>a</sup>

Present preferred career	Changing to teaching (%)	Changing to research (%)	No change (%)	N	Total %
College teaching	19	19	62	888	100
Research field	10	26	64	936	100
Other <sup>b</sup>	8	17	75	812	100

<sup>a</sup> At this point our discussion of changers is limited only to those who report a change toward research or teaching. A third group of ambiguous changers include those who show a preference for both teaching and research, either of these in conjunction with administration, or some other alternatives in addition to those mentioned above.

<sup>b</sup> The "other" group includes students who wish to enter a professional practice, administration, secondary school teaching, parish work, etc.

sample do report a change of some kind in their career preferences.

### RESULTS AND DISCUSSION

Table 1 shows which students in terms of present career preference have changed and the direction of that change.

From Table 1 it can be seen that about a third of the students who plan a career in either teaching or research indicate that there has been a career interest change of some kind. Among the "others" there is less overall change —amounting to one out of every four subjects in that category. In respect to direction of change Table 1 shows: (a) those who state a current career preference for teaching are evenly split—19% stating a stronger interest in teaching and 19% a stronger interest in research, (b) 10% of the researchers are now more favorably disposed towards teaching while 26% state they now find research more attractive as a postgraduate career, (c) in the "other" category change goes in the direction of research at a better than 2 to 1 rate. The data suggest that the graduate school experience has the effect of bringing students closer to a research orientation, and that this move toward research is felt by students in each of the occupational groups.

### Differences in Values and Attitudes

Included in the graduate student questionnaire is a set of "dilemmas," a group of imaginary situations in which the respondent is asked, after being confronted with two lines of action with alternate implications: "What would you do?" or "What would you prefer?"

One dilemma has to do with the distinction between "localism" vs. "cosmopolitanism" in academic careers and consists of the following question: "In the long run would you rather be known and respected (a) throughout the institutions where you work or (b) among specialists in your field in different institutions?" Subjects who checked *a* were classified as "locals" while those who checked *b* were considered "cosmopolitans."

Table 2 shows that in each case

TABLE 2  
CAREER PREFERENCE, CHANGE, AND  
"LOCAL-COSMOPOLITAN"

Present preferred career	Cosmopolitan (%)			
	No change in career preference	N	Change in career preference	N
Teaching	48	552	53	362
Research*	77	586	67	587

\*  $\chi^2$ —significant at <.001 level (research, changers, nonchangers).

TABLE 3  
CAREER PREFERENCE, CHANGE, AND  
WORK ORIENTATION

Present preferred career	Percentage who would devote all or the bulk of their uncommitted time to work			
	No change in career preference	N	Change in career preference	N
Teaching*	69	552	52	371
Research	69	554	64	590

\*  $\chi^2$ —significant at <.001 level (teachers—changers and nonchangers).

those with the research orientation show the greatest desire for the national recognition. The difference between the changers and the non-changers is found when the actual percentage differences are compared. The difference is one of intensity and not direction. Assuming that one's cosmopolitan orientation is tied to his career preferences, the results in Table 2 are what might be expected. Researchers who have not changed are most cosmopolitan as this is the type of professional view that would go with an unchanging attachment to a research career. Teachers who have not changed are least cosmopolitan since their frame of reference and operation would most likely be the local academic institution.

Significant differences were also found when respondents in each of the four groups were compared in respect to the *postgraduate* use of free time. Table 3 shows that while almost two-thirds of the research changers are inclined to devote the major part of their time to their work, only about half of the changers toward teaching are so inclined. A tentative interpretation would be that changers to teaching consist of a number of students who see teaching as a way of life in which the individual is able to regulate his time since he is guided by the num-

ber of hours a specific class meets. It may well be that these are graduate students who hold the more traditional (and perhaps naive) view of the academic world—one in which the professor prepares his lectures, meets his classes, and then retires to the solitude of his study. The nonchanging teachers appear to be no different than the researchers in the use which they would make of their free or uncommitted time. Career preference and degree of career commitment are also found to be related to how respondents view the teaching and research activity.

Table 4 shows that researchers are least likely to agree with the statement, "teaching is intrinsically more satisfying than research," those who have changed to teaching, and those teachers who have not changed being most in agreement with the statement. Not unlike the "local-cosmopolitan" situation it would appear that the responses of the changers to teaching and research is partially determined by the fact that these are students in transit. Included are: (a) students who have moved from teaching to research and now question the validity of the statement; (b) students who entered the graduate system with a preference for teaching and who have withheld to some extent the research influences of the graduate school, and

TABLE 4  
CAREER PREFERENCE, CHANGE, AND  
TEACHING VERSUS RESEARCH

Present career preference	Percentage agreeing that teaching is intrinsically more satisfying than research			
	No change in career preference	N	Change in career preference	N
Teaching	64	543	56	363
Research*	7	580	26	570

\*  $\chi^2$ —significant at <.001 level (researchers—changers and nonchangers).

in a sense have rededicated themselves to teaching; (c) those students who have made the shift from research to teaching. On the other hand, the non-changers present a different picture. Here are people who have retained that which they brought with them into the system in respect to teaching and research. The nonchanging researchers are strongest in their convictions that in fact research is more satisfying than teaching. The teachers who have not changed are most willing to accept the idea that teaching does hold an intrinsic satisfaction over research.

Responses to yet another "dilemma" situation provide further evidence to support the proposition that there are significant differences in attitudes and values when comparisons are made between teaching and research oriented students. In this case respondents were asked the following:

You are a full professor in a small but good department. Two men are being considered for a tenure position, but only one can be appointed.

*Smith* is very well liked by the students, faculty, and throughout the institution. He is an effective teacher and is extremely well read, both in his field and in other areas.

It is quite clear, however, that he is never going to produce research or work of scholarship which will make a fundamental contribution in the field.

*Jones* is known to everyone as a *prima donna*. He is difficult to get along with, and students complain about his lack of interest in teaching. It is quite clear, however, that in the long run he is destined to be one of the major intellectual figures in the field.

Which man would you vote for when the department makes its recommendations?

Each group, it would appear, is most eager to give support to his own kind. The teachers, be they changers or non-changers, are most enthusiastic about *Smith*. The researchers, especially

TABLE 5  
CAREER PREFERENCE, CHANGE, AND  
TENURE DILEMMA

Present career preference	Percentage who would choose Smith			
	No change in career preference	N	Change in career preference	N
Teaching	88	552	85	370
Research	68	593	71	588

those who report no change, are least interested in bringing *Smith* into the department. While Table 5 does support a general hypothesis of differences among teachers and researchers, it presents data of additional interest. While each group indicates a preference for a colleague who is most like themselves, it will be noted that in each cell the majority express a preference for *Smith*—the teacher. It is possible to only speculate why this is so. Perhaps individuals prefer *Smith* since he seems most congenial and easiest to get along with. On the other hand it may well be that researchers and teachers are expressing a need for something they find lacking in their own graduate experience, namely a faculty member who would give more of his time to his students and less to his research interests. \*

#### SUMMARY

The results of a national survey of 2,842 graduate students would indicate that graduate students do alter their career preferences as they progress through their professional training. Significant differences are found in attitudes and values when these students are compared in respect to career preference and degree of career commitment.

The data presented would suggest that a form of "anticipatory socialization" is in operation and has the effect

of preparing students for their post-graduate careers.

#### REFERENCES

- BECKER, H. B., & GEER, D. The fate of idealism in medical school. *Amer. soc. Rev.* 1958, **23**, 50-56.
- BECKER, H. B., & STRAUSS, A. Careers, personality and adult socialization. *Amer. J. Soc.*, 1956, **62**, 253-263.
- GOTTLIEB, D. J. Processes of socialization in American graduate schools. *Soc. Forces*, 1961, **40**, in press.
- JACOB, P. E. *Changing values in college*. New York: Harper, 1957.
- MERTON, R. K., & KITT, A. S. Contribu-  
tions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of the American soldier*. Glencoe: Free Press, 1955.
- MERTON, R. K., READER, G., & KENDALL, P. *The student physician: Introductory studies in the sociology of medical education*. Cambridge: Harvard University Press, 1957.
- SPAETH, J. L. Value orientations and career preferences of graduate students. Paper read at the Chicago meetings of the American Sociological Society, September 1959.

(Received February 8, 1961)

# THE ACQUISITION AND RETENTION OF PAIRED ASSOCIATES BY GOOD, AVERAGE, AND POOR READERS<sup>1</sup>

WAYNE OTTO

*University of Oregon*

The present study investigates the relative abilities of good, average, and poor readers to learn and to recall a list of paired associates. The particular variables selected for study in this connection were chosen primarily because of questions raised by an earlier, exploratory study.

The exploratory study employed a learning task similar to the one used in the present study: a list of five paired associates, each pair consisting of a common geometric form and a low association value trigram. The task discriminated between good and poor readers sufficiently well to suggest further experimentation in examining several specific questions raised by the study.

1. In the earlier study, the subjects' IQ scores varied considerably. With intelligence level differences minimized, would the demonstrated differences between good and poor readers' ability to master paired associates persist?

2. The exploratory study had employed good readers and extremely disabled readers. Would the differences persist if the poor reading group consisted of subjects reading in the lower quarter of their class? How would a group of "average" readers perform?

3. There appeared to be a tendency for visual-auditory reinforcement to

result in more rapid learning than did purely auditory reinforcement. With more adequate controls, would a particular mode of reinforcement continue to be more effective than another? Or, with provision made for the study of a possible interaction between mode of reinforcement and reading level, might one mode of reinforcement prove to be better for poor readers while another would be better for good readers?

4. If grade placement of subjects were to be varied systematically would this grade placement variable interact significantly with reading level and/or mode of reinforcement? For example, might not the most effective mode of reinforcement vary from grade level to grade level?

5. What is the effect of these variables—reading level, grade placement—chronological age, and mode of reinforcement—upon total trials required to relearn the list after 24 hours? The results of the earlier study suggest that there would be no difference in number of relearning trials required by good and poor readers.

Studies have been reported that have been concerned with one or more of the above questions. Five general statements regarding these studies can be made here.

1. Level of intelligence appears to be a significant variable in associative learning. There appears to be an inverse relationship between IQ and trials required to master a list of paired associates (Gates, 1930).

2. Studies show that associative

<sup>1</sup> The study is based upon data collected for a doctoral thesis at the University of Wisconsin. The writer is grateful for helpful comments to his thesis committee: Theo-dore L. Harris, Chester W. Harris, and Virgil E. Herrick.

learning ability and reading ability are directly related, but there is little agreement as to why or to what extent this is so (Fildes, 1921; Gates, 1935; Hincks, 1926; Monroe, 1932; Raymond, 1955; Vernon, 1958).

3. In general, studies have failed to demonstrate a clear advantage for any one type of reinforcement. The usual types of reinforcement investigated have been auditory, visual, and tactile/kinesthetic, presented singly and in combination (Berman, 1939; Burt & Lewis, 1946; Kirk, 1933; Raymond, 1955; Rivkind, 1958; Stauffer, 1948; Williams, 1952).

4. Studies have not shown adequate control of the learning tasks employed to permit conclusive interpretations of results. The failure of many studies to demonstrate advantages of any particular mode of reinforcement may simply reflect shortcomings of the task employed rather than absence of effects of the task variables.

5. There has been no real concern with what might be important interactions. Studies have generally been concerned with but one level of reading ability, with a particular mode of reinforcement, with a general range of chronological ages or levels of ability at the possible cost of what might have been significant interactions between two or more of these variables.

The present study, then, attempts to control and study what appear to be pertinent variables and make allowance for consideration of possible significant interactions between selected variables. The study is an experimental investigation of relationships that, once demonstrated, can serve as bases for further research in learning to read situations.

Three hypotheses provide a focus for analysis of the data and subsequent discussion.

1. Good readers, average readers,

and poor readers will differ in the number of trials required to learn a list of paired associates.

2. Mode of reinforcement will interact significantly with reading level and/or grade level.

3. Good, average, and poor readers will not differ in number of trials required to relearn a list of paired associates 24 hours after acquisition trials.

## METHOD

Three major variables were studied: reading level, grade placement-chronological age, and mode of reinforcement.

### Subjects

One hundred and eight subjects (four replications of the  $3 \times 3 \times 3$  design), chosen from a total population of about 2,900 elementary school children from three schools, served in the experiment. All subjects met the following criteria:

*Intelligence.* A group test IQ score between 95 and 110. Scores were from the Otis Quick Scoring at Grade 2, and from forms of the California at Grades 4 and 6. The attempt was to hold intellectual capacity within the average range by eliminating high and low deviant scores. The "average" score was subject to corroboration by the classroom teacher.

*Reading level.* Reading test scores falling in one of three specified categories denoted good, average, or poor. Scores were from forms of the Metropolitan Achievement Tests. Raw test scores were converted into stanine equivalents, which divide the distribution into nine subdivisions of approximately the same range. Thus, operationally defined, good readers were those whose test scores fell in the seventh, eighth, or ninth stanines; average readers had scores falling in the fifth stanine; and poor readers had scores falling in the first, second, or third stanine. The fourth and sixth stanines were dropped out to avoid overlap between groups. The level to which a subject was assigned was subject to corroboration by the classroom teacher.

*Grade level.* Placement at second, fourth, or sixth grade. Subjects were chosen from three grade levels in order to study the possible effect of educational development-chronological age upon efficiency in the learning task. Chronological age range was limited to 12 months to eliminate any wide

deviations from the usual age at grade level.

A pool of children who met the criteria was selected and four subjects were then assigned to each experimental condition. Since the possible effect of sex differences upon learning efficiency under any of the experimental conditions was not known, two boys and two girls were assigned to each experimental condition. Because no systematic differences were demonstrated on the basis of sex, this will not be further discussed.

### Task

*List.* The paired-associates list employed in the study evolved from the preliminary study done by the writer. Each pair consists of a common geometric form and a low association value CVC trigram from the Archer (1960) list.

The specific geometric forms were chosen because preliminary investigation showed that most elementary school children had a dominant name for each one (diamond, circle, triangle, star, square). Geometric form-trigram pairs appeared to be preferable to trigram-trigram pairs to avoid needlessly confusing the younger subjects and subjects with poor reading ability with two word-like stimuli in each pair.

The trigrams chosen had association values of 25% or less according to the Archer list. The five trigrams were chosen to employ the short sounds of the common vowels—*a, e, i, o, u*—without repeating any consonant in the list.

The actual list follows. (The common name of the form is given here in lieu of the form itself.) Diamond-sep, circle-miv, triangle-wuc, star-yad, square-gox. In those reinforcement conditions in which the subject received visual presentation of the trigrams, the trigrams were printed in lower case manuscript style.

Since it was felt that some basis for comparison of associations evoked by the individual forms and trigrams in the different groups of subjects would have value in interpreting the data, an association test was implemented in the study. This test and results will be alluded to, but space does not permit further elaboration here.

*Apparatus.* A Stone memory drum was employed in controlling presentation of forms. The original box encasing the revolving cylinder provided an opening  $5'' \times .5''$  for presentation of stimuli. A slide opening was devised to permit the experimenter to expose  $1'' \times .5''$  portions of the total stimulus sheet mounted on opening. The stimulus sheet mounted on

the cylinder was so arranged that five different orders of presentation were possible. The forms presented on the memory drum were  $\frac{3}{8}''$  high.

*Acquisition.* All testing was done in a room with only the experimenter and the subject present and with a minimum of distracting outside noise. The writer served as experimenter throughout the experiment.

Each form was presented for 4 seconds—at which point the subject was expected to say the form name—followed by a 4-second presentation of the form-trigram, followed by a 4-second rest. There was a 16-second intertrial rest. Pairs were first presented in serial order until the subject reached the criterion of one correct anticipation of form names for the entire list. Pairs were then presented in scrambled order to the same criterion. The total score was serial plus scrambled trials to criterion. This sequence of presentation was employed because immediate scrambled presentation proved to be disconcerting to the younger subjects. The scrambled presentation countered any sequential learning effect.

*Recall/relearning.* Subjects were asked to relearn the list after 24 hours. Pairs were presented in scrambled order in the same manner as in acquisition trials. The subjects continued to the criterion of one correct anticipation of form names for the entire list.

### Reinforcement

The term "mode of reinforcement" is used in the present paper to refer to the manner in which the correct form-trigram pairs were presented to the subjects. "Confirmation" might be an equally apt term. The actual procedure is described in the following paragraphs. The point to be made here is that the word "reinforcement" does not conform to the strictly conventional usage.

Three modes of reinforcement were employed. In practice the reinforcement was cumulative: auditory, visual-auditory, and kinesthetic-visual-auditory. For the sake of simplicity, however, the three modes of reinforcement will be denoted auditory, visual, and kinesthetic, respectively.

*Auditory.* The form was presented for 4 seconds, followed by a second 4-second presentation of the form accompanied by the experimenter's articulation of the appropriate trigram (form name).

*Visual.* The form was presented for 4 seconds, followed by a second 4-second pres-

TABLE 1  
MEAN ACQUISITION TRIALS

Grade			Reading level			Reinforcement		
2	4	6	Good	Average	Poor	Auditory	Visual	Kinesthetic
16.42	11.42	9.14	8.67	13.03	15.28	13.36	12.33	11.28

entation of the form accompanied by the experimenter's articulation of the appropriate trigram plus visual presentation of the trigram. For the visual presentation, the experimenter held a card bearing the printed trigram contiguous to, and to the right of, the form presented by the memory drum. Cards,  $1'' \times 2''$ , with  $\frac{1}{2}'' \times \frac{1}{2}''$  centered letters, approximately the size of the forms, printed in black were employed.

*Kinesthetic.* The form was presented for 4 seconds, followed by a second 4-second presentation of the form accompanied by the experimenter's articulation of the appropriate trigram plus a visual-kinesthetic presentation of the trigram. For the visual-kinesthetic presentation, the subject was presented a card bearing the form-trigram pair and asked to trace the trigram with his finger. Cards,  $10'' \times 4''$ , with  $1\frac{1}{4}'' \times 2''$  letters, a size which the subject could readily trace, printed in black were employed. Letters were lower case manuscript, of the type most familiar especially to the second grade subjects.

*Pretraining.* Each subject was presented a card bearing a sample pair of stimuli (heart-keb) before acquisition trials began. The subject gave his name for the heart-shaped figure and was told that he would be asked to learn a new name. The test forms were then presented and paired with the appropriate trigrams in the manner of the reinforcement to be given. Subjects to receive kinesthetic reinforcement practiced tracing the sample trigram quickly.

#### RESULTS AND DISCUSSION

A comparison of mean IQ scores shows no significant differences between groups compared in subsequent analyses. It appears that the random assignment of the pooled subjects to experimental conditions was reasonably successful in eliminating between group IQ differences; therefore, between group differences do not appear

to be attributable to IQ differences between groups.

*Acquisition.* Mean acquisition scores for the different groups for each of the three main variables are shown in Table 1. An analysis of variance (Edwards, 1950) of the total acquisition scores is presented in Table 2. There was little evidence of heterogeneity of variance ( $\chi^2 = 30.72$  for  $26 df$ ,  $p > .20$ ). Each of the three main effects is significant at the .01 level. That differences in performance by grade level are dependent upon reading level of subjects is indicated by the significant grade  $\times$  reading level interaction. That differences due to mode of reinforcement are dependent upon grade placement of subjects is indicated by the grade  $\times$  reinforcement interaction. There is no significant reading level  $\times$  reinforcement interaction, nor is there a significant triple interaction.

Table 3 presents results of Tukey gap tests (Tukey, 1949), which were

TABLE 2  
ANALYSIS OF VARIANCE OF THE  
ACQUISITION TRIALS

Source	df	MS	F
Grade (G)	2	498.94	79.91**
Reading Level (L)	2	406.735	65.14**
Reinforcement (R)	2	39.065	6.25**
G $\times$ L	4	34.347	5.50**
G $\times$ R	4	30.975	4.90**
L $\times$ R	4	5.16	
G $\times$ L $\times$ R	8	11.562	1.85
Within	81	6.244	

\*\*  $p < .01$ .

TABLE 3  
RESULTS OF TUKEY TEST OF SIGNIFICANT GAPS

Grade	Grade		Reading level			Reinforcement		
	Mean trials	Gap	Level	Mean trials	Gap	Mode	Mean trials	Gap
2	16.42		P	15.28		A	13.36	
4	11.42	5.00**	A	13.03	2.25**	V	12.33	1.03*
6	9.14	2.28**	G	8.67	4.36**	K	11.28***	1.05*

\*  $p < .10$ .

\*\*  $p < .01$ .

\*\*\* The auditory/kinesthetic between means gap is significant ( $p < .01$ ).

run to determine where significant differences between group means existed within the three main effects. Significant differences in mean acquisition trials were demonstrated between each of the three grade levels and between each of the three reading levels. The differences between the mode of reinforcement mean acquisition trials are approaching the .05 level of significance for each mode of reinforcement, but only the difference between the auditory and visual mode of reinforcement meets the  $<.05$  requirement. Further interpretation must, of course, consider the significant interactions that have been demonstrated.

Several of the questions and hypotheses can be examined in view of the acquisition data.

Hypothesis 1: Good, average, and poor readers will differ in the number of trials required to learn a list of paired associates.

The hypothesis is accepted, with the restriction that the grade level of subjects tested will tend to influence the magnitude of differences demonstrated. Table 4 makes clear the extent of the influence of grade level upon differences between reading levels. Examination of the mean acquisition trials presented shows that poor readers in Grade 2 took almost twice as many trials to master the list as did good readers; whereas poor readers in Grade 6 took only 1.6 times as many trials as did good readers. Also, good readers at the three grade levels perform more nearly alike than do the poor readers. It is suggested that the task is not sufficiently difficult at the higher grade levels to preserve the marked dichotomy displayed at Grade 2.

Hypothesis 2: Mode of reinforcement will interact significantly with reading level and/or grade level.

The hypothesis is accepted: there is a significant grade  $\times$  reinforcement interaction. Mode of reinforcement does not, however, interact with reading level, contrary to what seems to be a basic assumption when added cues are stressed in much remedial work.

The nature of the grade  $\times$  reinforcement interaction is made clear by Table 5. It can be seen that the most

TABLE 4  
MEAN ACQUISITION TRIALS:  
GRADE  $\times$  READING LEVEL

Grade	Reading level		
	Good	Average	Poor
2	10.58	18.17	20.50
4	8.33	11.83	14.08
6	7.08	9.08	11.25

TABLE 5  
MEAN ACQUISITION TRIALS:  
GRADE X REINFORCEMENT

Grade	Reinforcement		
	Auditory	Visual	Kinesthetic
2	16.33	18.42	14.50
4	13.58	9.83	10.83
6	10.17	8.75	8.50

efficient mode of reinforcement (requiring the fewest trials) varies by grade level. With the second grade subjects, kinesthetic reinforcement appears to make for the fewest trials; whereas, visual reinforcement appears to be most efficient for the fourth grade subjects and kinesthetic and visual are almost equally effective for the sixth grade subjects. The implication seems to be that the function of different cues supplied by the several modes of reinforcement changes as subjects progress in grade level and chronological age. This may help to explain why gross studies often fail to show a particular mode of reinforcement to be better than others. In order to secure positive results in any predictable manner, one would need to control grade level-chronological age.

The answers to two of the questions raised by the earlier study are clear from the consideration of Hypothesis 1 and Hypothesis 2: (a) When intelligence is held within a stated range, will good, average, and poor readers differ in the number of trials required to learn a list of paired associates?

(b) Will there be significant interactions between reading level, mode of reinforcement, and grade level? The answers to both questions are affirmative, subject to the restrictions pointed out in connection with the hypotheses.

The answers to two other questions are also affirmative: (a) Will the traditional modes of reinforcement significantly influence rate of learning? It has been shown that the particular mode of reinforcement that is most effective will vary from grade to grade. (b) Is grade level-chronological age a significant variable in determining the number of trials required to master a list of paired associates? The affirmative answer is restricted by the fact that differences in number of trials required at different grade levels will be influenced by the reading proficiency of the subjects and by the mode of reinforcement employed.

*Recall and relearning.* "Recall score" is the number of form names correctly anticipated 24 hours after the acquisition trials. "Relearning score" is the total number of trials required to enable the subject to correctly anticipate form names for the entire list presented in scrambled order.

Mean recall and relearning scores for the several groups by variable are presented in Table 6. It can be seen that in general the mean recall scores vary in the same manner as the mean relearning scores. For this reason and because relearning efficiency probably

TABLE 6  
MEAN RECALL AND RELEARNING SCORES

	READING AND RELEARNING SCORES								
	Grade			Reading level			Reinforcement		
	2	4	6	G	A	P	V	A	K
Recall	2.86	3.17	3.39	3.56	2.72	3.14	3.17	3.0	3.25
Relearning	2.67	2.11	2.11	2.08	2.50	2.31	2.44	2.22	2.22

most adequately reflects retention, only the relearning scores were subjected to analysis of variance. This analysis is presented in Table 7. Only grade level was found to be significant. Table 6 shows that the mean relearning scores for Grades 4 and 6 are identical. Consequently, it can be concluded that the only significant differences in relearning scores are attributable to the fact that second grade subjects took more trials to relearn than did fourth and sixth grade subjects.

Hypothesis 3 and the related question can be examined in view of the relearning data.

**Hypothesis 3:** Good, average, and poor readers will not differ in the number of trials required to relearn a list of paired associates 24 hours after acquisition trials. And the question: What is the effect of the variables—reading level, grade placement, and mode of reinforcement—upon total trials required to relearn the list after 24 hours?

The hypothesis is accepted: there is no difference in relearning performance due to reading level of the subjects. The answer to the question is that grade placement of the subjects alone significantly effects relearning after 24 hours.

That there were no differences between good, average, and poor readers in trials required to relearn the list seems to be a significant finding. This suggests that given sufficient reinforcement the future performance of good, average, and poor readers would tend to be similar. Thus, while subjects might require different numbers of trials to master a task, they might be expected to attain about the same performance level. This is, of course, part of the rationale proposed by some for the use of "teaching machines" or "autoinstructional devices."

TABLE 7  
ANALYSIS OF VARIANCE OF  
RELEARNING TRIALS

Source	df	MS	F
Grade (G)	2	3.70	5.69**
Level (L)	2	1.57	2.42
Reinforcement (R)	2	.97	1.49
G × L	4	.46	
G × R	4	.13	
L × R	4	.24	
G × L × R	8	.98	1.51
Within	81	.65	

\*\*  $p < .01$ .

*Association test.* The association test results indicated that the poor readers gave more associations to the individual stimuli of the paired-associates list than did the good readers. As pointed out, space limitations preclude discussion of the association test; but the implication appears to be that there is no clear-cut relationship between the number of associations evoked (fluency) and efficiency in the present learning task.

#### SUMMARY AND CONCLUSIONS

A  $3 \times 3 \times 3$  factorial design was employed to study the effects of three levels of reading proficiency, three levels of grade placement (chronological age), and three modes of reinforcement upon acquisition of a list of paired associates. One hundred and eight children whose IQ scores were in the average, 95–110, range provided data for the investigation.

The results warrant the following conclusions:

1. Good, average, and poor readers—in that order—require increasingly more trials to master a list of paired associates.

2. Sixth, fourth, and second graders—in that order—require an increasing number of trials to master a list of paired associates; but these differences

are significantly influenced by both the reading proficiency of the subjects involved and the mode of reinforcement employed.

3. Mode of reinforcement as a variable interacts significantly with grade level. Therefore, studies purporting to investigate the relative efficiency of different modes of reinforcement must recognize grade level of the subjects as an effective variable.

4. There is not a significant interaction between mode of reinforcement and reading level. This is contrary to the assumption in instructional practice that additional cues will be of value in a remedial teaching/learning situation.

5. Analysis of total relearning scores showed grade level to be the only significant variable. Thus, it appears that once they have mastered a list of paired associates, poor readers will retain the associations as well as good readers.

6. Analysis of the results of the association test indicates that the differences in ability to learn a list of paired associates demonstrated by good, average, and poor readers cannot be explained simply in terms of number of meanings or associations attached to the stimuli.

Two limitations seem obvious in considering the significance of the present study for the teaching/learning of reading: (a) The universality of the present results needs to be demonstrated. (b) The functional relationship of the present task to learning to read is not known. It is felt, however, that the present results can be of value in formulating hypotheses to be evaluated when reading tasks are employed in the experimental setup.

## REFERENCES

- ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, 74(10, Whole No. 497).
- BERMAN, A. The influence of the kinesthetic factor in the perception of symbols in partial reading disability. *J. educ. Psychol.*, 1939, 30, 187-198.
- BURT, C., & LEWIS, R. B., Teaching backward readers. *Brit. J. educ. Psychol.*, 1946, 16, 116-132.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- FILDES, LUCY G. A psychological inquiry into the nature of the condition known as congenital word blindness. *Brain*, 1921, 44, 286-307.
- GATES, A. I. *Interest and ability in reading*. New York: Macmillan, 1930.
- GATES, A. I. *The improvement of reading*. (Rev. ed.) New York: Macmillan, 1935.
- HINCKS, E. M. *Disability in reading and its relation to personality*. Cambridge: Harvard Univer. Press, 1926.
- KIRK, S. A. The influence of manual tracing on the learning of simple words in the case of subnormal boys. *J. educ. Psychol.*, 1933, 24, 525-535.
- MONROE, MARION. *Children who cannot read*. Chicago: Univer. Chicago Press, 1932.
- RAYMOND, DOROTHY M. The performance of reading achievers on memory span and associative learning tests. *J. educ. Res.*, 1955, 48, 455-465.
- RIVKIND, H. C. The development of a group technique in teaching word recognition to determine which of four methods is most effective with individual children. Unpublished doctoral dissertation, University of Florida, 1958.
- STAUFFER, R. G. Certain psychological manifestations of retarded readers. *J. educ. Res.*, 1948, 41, 436-452.
- TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, 5, 99.
- VERNON, M. D. *Backwardness in reading*. Cambridge: Cambridge Univer. Press, 1958.
- WILLIAMS, GERTRUDE H. An investigation of the relative values of certain word-learning techniques for retarded readers. Unpublished doctoral dissertation, Temple University, 1952.

(Received March 8, 1961)

## THE CONTRIBUTION OF RESPONSE IN MENTAL MEASUREMENT

WILLAVENE WOLF

AND

JAMES B. STROUD

*Ohio State University*

*State University of Iowa*

Around the turn of the century, psychologists were active in investigating the relationship between speed of response and criteria of intelligence, as seen, for example, in the work of Gilbert (1894) and Wissler (1901). Most of the early workers dealt with simple motor responses, as in reaction time studies, or at best with simple responses. There appears to be little support for the existence of a general speed of response factor, as Tinker (1931) has pointed out. Generally negligible relationships have been obtained between speed of response on simple tasks and measures of intelligence. When significant relationships have been obtained, they have generally been interpreted as a function of the character of the specific task employed.

Thus it appears that rate of response on a speed test—one on which rate of response is the main determinant of the score—is essentially unrelated to performance on measures of intelligence. It does not necessarily follow, however, that rate of response on a power test is unrelated to criteria of intelligence. A subject who successfully completes difficult intellectual tasks quickly may have more mental ability than one who completes them, but must do so slowly. In such situations, it is reasonable to suppose that power makes for speed, as Kuhlmann (1939) and Tinker (1931) argued.

In addition to a concern about this general problem, the writers were interested in two intelligence tests: Raven's Progressive Matrices (revised order, 1956) and the Leiter International Performance Scale—both non-

verbal tests. Specifically, we were interested in investigating the relationships between rate of response to the individual items on these two tests and the total score (i.e., number right on the two tests, respectively), and certain independent measures. Some recent work with Progressive Matrices yielded rather low correlations between total scores and achievement measures, around .40 (Knief & Stroud, 1959). The form of the test and the method of administration make this an ideal test for use with the deaf and others handicapped in the use of a language. Obviously any improvement in its predictive power would be to the good. Some items from the Leiter were selected and used simply to further the interest in the general problem—the contribution of rate scores on a power test.

### PROCEDURE

The subjects in this investigation were 102 fifth grade pupils in two small midwestern communities, representing the entire fifth grade enrollment in the two communities. To these pupils were administered the Progressive Matrices, one-half of the tests in the Leiter scale, the Iowa Tests of Basic Skills, and the Lorge-Thorndike Intelligence test, Verbal Battery, Level 3. The Leiter tests used, starting with Year VI, were as follows: VI, 2 and 3; VII, 2 and 4; VIII, 2 and 4; IX, 3 and 4; X, 2 and 3; XI, 1 and 4; XII, 2 and 4; XIII, 1 and 4; XIV, 1 and 3; XV, 1 and 4; XVI, 3 and 4; XVII, 2 and 3. Progressive Matrices and the Leiter scale were administered individually. The examiner timed the

response to each item by means of a stop watch. The only alteration made in the directions was the addition of the sentence, "Do the items as quickly as you can, trying to get all the answers right."

The mean Lorge-Thorndike IQ of the sample was 109. The *SD* was 14. The mean PM raw score was 27.85. The *SD* was 8.19. The distribution was approximately normal.

In addition to the number right score on Progressive Matrices and the mental age score on the Leiter scale, a time score, based upon the average time required to compile the items done correctly, was computed for each pupil. The MA value of each item used on the Leiter scale was doubled, inasmuch as only one-half of the items was used. Progressive Matrices is made up of items of ascending difficulty. Thus it seemed probable that the time scores on some parts of the scale might be more significant than those on other parts. The scale was divided into fourths, numerically, and average time scores were computed for each pupil on each fourth of the scale, making it possible to determine the relation between speed of response and the criterion measures for each fourth of the scale.

Time allowances on the Iowa Tests of Basic Skills are rather liberal, but the test does impose time limits. The number of subjects finishing each subtest was ascertained. When fewer than 90% of the subjects had finished a subtest, questions at the end were eliminated—back to a point reached by 90%.

### RESULTS

As a first step in the analyses of the data, zero order correlations were computed (*a*) between average time score on the PM and Leiter tests and ITBS and Lorge-Thorndike scores and

(*b*) between number right scores on the PM and Leiter (MAs) and ITBS and Lorge-Thorndike scores. The results are reported in Table 1. The correlations between Progressive Matrices and the Leiter scale on the one hand and ITBS and the Lorge-Thorndike test, on the other, are surprisingly low. Since in this investigation the two tests were not administered in the standard way, the coefficients reported in Table 1 should not be taken as representative of the predictive power of the tests.

In order to determine the effectiveness of combining the average time scores with the number right scores on the PM and Leiter tests in the prediction of the composite achievement scores and verbal intelligence test scores, multiple correlation procedures were utilized. Table 1 gives the multiple correlation coefficients for the variables involved. The composite ITBS scores and the Lorge-Thorndike test scores are regarded as the dependent variables, and average time scores and number right scores on the PM and Leiter as the independent variables. Beta weights were computed and tested for significance. They are also reported in Table 1. Combining, in multiple correlation, time and number right scores on the PM and Leiter tests increased these tests' prediction on achievement and verbal intelligence scores to about the same degree. All beta weights, except that involving the Leiter time and the Lorge-Thorndike scores, were significant at the .01 level. Thus, the addition of time scores to PM and Leiter number right scores may be considered to have increased significantly the correlations between these two tests and the ITBS and Lorge-Thorndike scores.

The multiple correlation analysis indicated that making time a part of the scores increased the predictive

value of both the PM and Leiter tests. While the Leiter test did not lend itself to further analysis, since the subtests attempted by the respective subjects varied, further analysis was undertaken on the Progressive Matrices. To determine whether or not time on any fourth of the PM was more important for predictive purposes than was time on any other fourth, the PM was divided into sets of 14 or 15 items and the average time scores on each set were correlated with the ITBS and Lorge-Thorndike scores. The fourths represent roughly four different levels of difficulty. Only 74 subjects earned scores on each of the four sets of items. Thus this analysis was based on 74 cases. Table 2 presents the zero order correlations obtained between the average time scores on each set of items and the ITBS and Lorge-Thorndike scores. The only significant correlation was that between average time for the third set of items on the PM and the Lorge-Thorndike test.

Multiple correlations were computed to determine whether or not time scores on any fourth of the items on the PM when combined with the total number right scores yielded higher *R*s than those of any other

TABLE 2  
ZERO ORDER CORRELATIONS BETWEEN AVERAGE TIME SCORES FOR EACH FOURTH OF THE PROGRESSIVE MATRICES AND ITBS, LORGE-TORNDIKE, AND TOTAL NUMBER RIGHT ON PM  
(*N* = 74)

Average time	ITBS	Lorge-Thorndike
All correct items on the first fourth of Progressive Matrices	-.175	-.232
All correct items on the second fourth of Progressive Matrices	-.178	-.147
All correct items on the third fourth of Progressive Matrices	-.231	-.258*
All correct items on the last fourth of Progressive Matrices	-.104	-.030

Note.—Raven Time on first 14 items vs. Raven Right,  $r = .157$ ; Raven Time on second 15 items vs. Raven Right,  $r = .230$ ; Raven Time on third 15 items vs. Raven Right,  $r = -.458^{**}$ ; Raven Time on fourth 15 items vs. Raven Right,  $r = -.298^*$ .

\* Significant at .05 level.

\*\* Significant at .01 level.

fourth. The results are reported in Table 3. Beta weights were computed and tested for significance. They are also reported in Table 3.

The *R*s involving the time scores of the third fourth of items are higher than those involving any other fourth. One might have predicted that the magnitude of the *R*s would have increased with the difficulty of the items; i.e., that those of each succeeding fourth of the items would have been greater than those of the preceding fourth. Indeed a tendency in this direction was evidenced for the first three fourths. It seemed likely that the items comprising the last fourth of the test would, for these fifth grade pupils, be quite difficult and that, therefore, the average time scores of the various pupils would be based upon relatively few correctly

TABLE 1  
CORRELATIONS BETWEEN THE RAVEN AND LEITER TESTS AND ITBS AND LORGE-TORNDIKE TEST AND BETA WEIGHTS

Variable	ITBS	$\beta$	L-T	$\beta$
Raven Time	-.174	-.35**	-.178	-.37**
Raven Right	.217*	.38**	.254**	.43**
Multiple Leiter	.380**		.420**	
Leiter Time	.060	-.36**	.078	-.043
Leiter MA	.291**	.56**	.357**	.680**
Multiple	.310**		.460**	

Note.—Raven Time vs. Raven Right,  $r = -.465$ ; Leiter Time vs. Leiter Right,  $r = -.750$ .

\* Significant at .05 level.

\*\* Significant at .01 level.

TABLE 3  
MULTIPLE CORRELATIONS INVOLVING EACH  
FOURTH OF PM ITEMS

Variable	<i>R</i>	$\beta$	$\beta$
6.15	.30*	1 = -.21	5 = .19
6.25	.37**	2 = -.34**	5 = .36**
6.35	.45**	3 = -.46**	5 = .45**
6.45	.33*	4 = -.30	5 = .37**
7.15	.37**	1 = -.28*	5 = .30**
7.25	.38**	2 = -.32*	5 = .40**
7.35	.52**	3 = -.52**	5 = .52**
7.45	.32*	4 = -.22	5 = .37**

Note.—1. Average Time on first 14 items on Raven's PM.  
 2. Average Time on second 15 items on Raven's PM.  
 3. Average Time on third 15 items on Raven's PM.  
 4. Average Time on fourth 15 items on Raven's PM.  
 5. Number right on complete Raven's PM.  
 6. Iowa Tests of Basic Skills.  
 7. Lorge-Thorndike Intelligence Test.

\* Significant at .05 level.  
 \*\* Significant at .01 level.

done items. This, if true, should tend to lower the reliability of the average time scores for this fourth of the test. Accordingly, the number of items successfully completed in each fourth of the test, and upon which the average times were based, was ascertained. These, for the successive fourths, were found to be 12.7, 7.7, 6.5, and 2.1. Thus there is room for doubt that the apparent exception to the trend for the *r*'s to increase in magnitude seen at the last fourth of the test is genuine. When those pupils who got only one right or none, were eliminated, the mean number right was still only 2.8 on the last fourth of the test.

One additional treatment of the data for Progressive Matrices was undertaken. Median time scores were computed for each item and one point was given for each time score which was faster than the median score. These points for time were then added to the number right score on the PM for each pupil and a new score obtained. Scores

thus derived were then correlated with the criterion measures. Coefficients of .35 and .39 were obtained between the new PM scores and ITBS and Lorge-Thorndike scores, respectively, as compared with .217 and .254 between the PM number right scores and the same two criterion measures.

The data suggests that speed of response when combined with number right on the PM and Leiter increases significantly the correlation between these tests and the ITBS and Lorge-Thorndike scores.

In one recent investigation an average *r* of .41 was obtained between Progressive Matrices scores and various ITBS subtest scores, on 164 fourth grade pupils (Knief & Stroud, 1959). Most reported correlations that we have seen run somewhat higher than this. In the present investigation (*N* = 102), an *r* of .217 was obtained between PM scores (number right) and ITBS composite scores and an *r* of .254 between PM scores and Lorge-Thorndike verbal scores. Leiter MAs (based upon one-half of the items) correlated with these same two criterion measures to the extent of .291 and .357. When multiple *R*s were computed between (a) PM number right and average time scores and (b) Leiter MA and average time scores and these same two criterion measures, values of .380 and .420 were obtained for PM and of .310 and .460 for the Leiter. A coefficient of -.465 was obtained between average time values and number right scores on PM, and of -.75 between Raven average time values and MAs.

What the results of this investigation tend to show is that the speed with which correct responses are made on the PM and Leiter tests, when combined with the conventional scores, makes a significant contribution to the predictive power of these two tests, as

signified by increased correlations with the criterion measures. It is true, of course, that the correlations are still quite low, but the investigation does point up the importance of speed of response in the measurement of intelligence. Thus the results support the proposition that a person who can perform mental feats, especially difficult ones, quickly should be judged to have more ability than one who does so slowly.

#### SUMMARY

To investigate speed of response as a factor in mental measurement, the Progressive Matrices and the Leiter International Performance Scale were administered in individual sessions to 102 fifth grade pupils, and a time score was determined for each pupil.

The utilization of the time values together with the conventional scores resulted in correlations between the two tests in question (PM and Leiter) and the two criterion tests, the Iowa Tests of Basic Skills and the Lorge-Thorndike, significantly higher than the correlations obtained between the

conventional scores alone and the two criterion tests. Further analysis of the PM data suggested that the contribution of the time values increased from fourth to fourth of the items, and thus with difficulty, through the third fourth.

The results imply that the person who is able to deal more quickly with intelligence test items has more ability than the slower, but equally accurate individual.

#### REFERENCES

- GILBERT, J. A. Researches on the mental and physical development of school children. *Stud. Yale Psychol. Lab.*, 1894, **2**, 40-100.
- KNIEF, L. M., & STROUD, J. B. Intercorrelations among various intelligence, achievement, and social class scores. *J. educ. Psychol.*, 1959, **50**, 117-120.
- KUHLMANN, F. *Tests of mental development*. Minneapolis: Educational Test Bureau, 1939.
- TINKER, M. A. The significance of speed in test response. *Psychol. Rev.*, 1931, **38**, 450-454.
- WISSLER, C. The correlation of mental and physical tests. *Psychol. Monogr.*, 1901, **3**(6, Whole No. 16).

(Received March 15, 1961)

## EFFECTS OF TRAINING AND ANXIETY UPON TEACHERS' PREFERENCES FOR INFORMATION ABOUT STUDENTS<sup>1</sup>

PHILIP J. RUNKEL AND DORA E. DAMRIN

*Bureau of Educational Research, University of Illinois*

A characteristic of the individual's perceptual or cognitive space which captures much attention in current literature—e.g., Bieri (1955), Bieri and Blacker (1956), Leventhal (1957), Lundy and Berkowitz (1957)—is the "complexity-simplicity" of the frame of reference within which the individual perceives environmental objects. Complexity-simplicity can be conceived as the dimensionality of the cognitive space. By dimensionality we do not refer to the nature or content of the dimensions mediating the response, but only to the sheer number of the dimensions used by the individual in discriminating environmental objects. This paper examines two hypotheses: (a) that training is related to cognitive dimensionality, and (b) so also is anxiety. The tests of these hypotheses were set in the context of teachers' thinking about problems of students.

### *Dimensionality and Knowledge*

Let us consider a couple of common notions about training or knowledge. On the one hand, it is widely held that training serves to rule out irrelevancies. By focusing attention upon what is relevant and important, training restricts the number of criteria by which matters are judged. On the other hand, a proposition dear to educators is that training is *liberating*; it provides the individual with an enhanced repertoire of criteria from which he can

select the most appropriate to guide his judgment. Whereas the untrained individual is more prone to adopt an irrelevant dimension or one given by his mood of the moment, the individual with sufficient training possesses a kit bag of relevant dimensions and can judiciously select from it dimensions suitable in different situations. From this point of view training increases the number of dimensions which, at any given moment, the individual can draw upon.

These two ideas are not contradictory. The first says that the untrained person's criteria or dimensions of judgment depend on his mood of the moment or upon surface irrelevancies in the situation, so that in meeting a series of situations he utilizes a large number of dimensions. Although he might react on the basis only of one or very few dimensions in any one situation, his dimensions over many situations would be numerous. The second phase begins when the individual acquires a little knowledge. He then looks for certain relevant and important characteristics no matter what the situation, so that the tally of dimensions he uses over a series of situations is small. Then with further training, the person comes to recognize more criteria as relevant to each situation, selecting certain ones as more important in some situations and others as more important in others; consequently the dimensionality of his judgments over situations increases.

These ideas would lead one to predict that the number of dimensions likely to be used in a single situation

<sup>1</sup>The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

would increase with training. Over a series of situations, however, dimensionality would first decrease and then increase as knowledge was acquired. The particular prediction made in terms of our data will be better understood if we first describe our instruments and method.

### METHOD

Teachers in 28 secondary schools in Illinois were given a questionnaire describing twelve imaginary problems which a teacher might encounter with students. For each problem, the teacher was asked to express his preferences among six types of information which he might seek in order to deal effectively with the problem. Problem 1, along with a condensation of the instructions given to respondents, looked like this:

#### *Problem 1*

Joan McCauley has not turned in a single assignment since the beginning of the year. You have warned her that failure to turn in her homework will result in her failing the course, but this has no effect. She sits and daydreams most of the time she is in class.

As a working guess, where would you say the source of this difficulty might lie? To check up on your working guess, what additional kinds of information would you seek? Please number, in order, the types of information (no more than three).

Achievement in school subjects	Family and home life	Intelligence	Fears and worries	Interests and ambitions	Personality and char- acter
--------------------------------------	-------------------------	--------------	----------------------	----------------------------	-----------------------------------

To illustrate further, three more of the problems follow:

Virginia Smith's parents want their daughter to go to college, and Virginia herself wants very much to go. However, she has no idea what to major in.

Jim has been sent to the principal's office by two different teachers for his rude and sarcastic behavior in their classes. His view of every instance is, "The teacher started it."

Glen's teachers report that the boy comes to school so tired every day that he is always falling asleep in their classes.

The school physician reports that he is basically in good health.

Each teacher chose three of the types of information and ordered them 1-2-3 according to his preference, and he did this 12 times, once for each problem. Using the 12 rankings given by a teacher (one ranking of three things for each of the 12 problems), it was possible to attempt to "unfold" them into one dimension according to the technique of Coombs (1953).

Unidimensional unfolding is a method for ascertaining whether an obtained rank order (or a set of rank orders) might have been generated from an underlying order. For example, suppose that a man with a wife and two children finds that six houses are available on the market, one with a combination living and sleeping room, another with one bedroom, another with two bedrooms, and so on up to five bedrooms. He might most prefer the house with three bedrooms, next the one with four, and thirdly the one with five. When he becomes the father of a third child, he might then most prefer the house with four bedrooms, then the one with five, and thirdly the one with three bedrooms. Both these rankings of three of the houses out of six could have been generated from the underlying dimension of the number of bedrooms in the houses. But if a man most preferred the house with, say, two bedrooms, he could not (on the basis of the number of bedrooms alone) next prefer the house with five—his next choice would have to be either the house with three bedrooms or the house with one. Thus some responses in rank order form are "unfoldable" into the number of bedrooms dimension and some are not. A more extended discussion of unfolding can be found in Coombs (1953).

An interpretation of unfolding in terms of the materials used in the present study might be illustrated as follows. Suppose that a teacher judged student problems in terms of what we might loosely call a "hard-soft" dimension. For one problem the teacher might feel that very "hard" information was most suitable, choosing first intelligence, then achievement, then interests, then family life, etc. On another problem the teacher might feel that an intermediate type of information was most suitable, choosing first interests, then achievement, then family life, then intelligence, etc. Both these preference orders could unfold into the same dimension, as will be seen below.

The unfolding technique does not, of course, depend on guessing what basis the

TABLE 1  
ILLUSTRATION OF HOW THE RESPONSES OF  
ONE SUBJECT TO THE 12 PROBLEMS  
FITTED INTO AN UNFOLDING SET

Type of information	Possible preference orders (each column shows the response to some problem)
Intelligence	1 2 3
Interests and ambitions	2 1 1 1 2 3
Achievement in school subjects	3 3 2 2 1 1 1 2 3 3
Personality and character	3 3 2 2 1 1 2 2 3 3
Family and home life	3 3 2 1 1 1 2
Fears and worries	3 2 1
Number of problems to which subject gave re- sponse shown in column <sup>a</sup>	1 2    1 1    1 1    2

<sup>a</sup> Responses of the subject which would not unfold with the other nine in this set were: Interests-family-fears, family-interests-personality, and fears-intelligence-interests.

subject might be using in organizing his preferences. The technique is entirely objective. As an illustration, the responses of one of the subjects are shown in Table 1. The 1-2-3 entries in the body of the table show one set of preference-orders which could be generated by the "underlying" order of information types listed at the left. For any given underlying order of the information types, and under the instruction to order three out of the six, there are eight sets (the table shows only one) of 13 response patterns, each being a set which could be generated by the underlying order when the respondent chooses, in order, the types nearest his "ideal" type. Eight sets are possible because of the effects of relative distances among the types; this, as well as other details of the unfolding technique, is explained by Coombs (1953).

The response patterns given by our illustrative subject are indicated at the bottom of the table. The particular unfolding set represented in the table accounted for nine of the subject's responses, and this was the most that any unfolding set would accommodate.

If any one of a teacher's 12 rankings failed to unfold with his others, we knew that at least two dimensions were required to encompass the 12 responses. We then went on to assume that the fewer of the teacher's 12 responses that would unfold together, the higher his dimensionality was likely to be. This would not necessarily be

true, of course. But we assumed that it was more likely to be true than not.

The maximum number of rankings unfoldable into a single dimension was taken directly as the individual's dimensionality "score." Since there were 12 problems, the dimensionality score could range from 1 to 12. A large number of unfoldable responses means low dimensionality, and vice versa. Actually, the obtained scores ranged from 3 (high dimensionality) to 10 (low dimensionality). If a teacher gave an unusable response to more than one problem, he was recoded NA and not given a dimensionality score.

The unfolding technique thus gave us a rough index of whether the teacher was using relatively few or relatively more dimensions in thinking about the types of information he would need in dealing with the problems described to him.

As a measure of training influences, we used a 60-item test developed in the course of our larger study: a test of Knowledge and Interpretation of Tests (KIT). We did not want a measure of years of experience in dealing with information about students, since, as it is often phrased, 10 years' experience often consists of 1 year's experience repeated 10 times over. Nor did we want a direct measure of amount of exposure to training, since there are great individual differences in the degree to which training "takes," especially when assessed a year or more after the last exposure to training. Consequently, we wanted what might be called a measure of the residual effects of training.

Some of the types of information offered the respondent were types typically obtained through the use of tests. This circumstance, together with the fact that training in the use of information about students usually includes training in constructing and interpreting tests, persuaded us that for the purposes of our hypothesis a measure of knowledge about testing would suffice to index the extent to which the individual's training would affect his tendencies in the use of information about students. Consequently, scores on the KIT were distributed against the dimensionality scores to test the hypothesis that increased training is associated with different dimensionalities. Some representative items from the KIT are given below.

In counseling junior high school students a teacher should be *most* cautious in her interpretation of the information she obtains from:

- A. vocational interest tests
- B. study skills tests
- C. reading comprehension tests
- D. performance tests

A student scored at the 75th percentile on a standardized achievement test. This means that:

- A. 75% of the norm group scored lower than he did
- B. 75% of the norm group scored higher than he did
- C. he answered 75% of the questions correctly
- D. he is in the upper 25% of his own high school class

On a group intelligence test, which pupil would be most apt to receive an IQ that is much lower than his true ability?

- A. Mary, whose parents speak only Spanish
- B. John, who failed in all of his school subjects last year
- C. Jane, who comes from a lower middle class home
- D. Bill, who spends all of his time in athletics instead of on his school work

The question arises what *kind* of relation should be predicted between our two measures—dimensionality (number of mutually unfoldable rankings) and training (score on the KIT). Can we specify precisely where, within the ranges covered by our measures, dimensionality should increase or decrease? We mentioned earlier that the predictions would be different if we were looking within a single situation or if we were looking at a series of situations. Was answering the Problems Questionnaire a single situation for our subjects, or was it a series of 12 situations? We could hardly know in advance.

Another difficulty in choosing a prediction was the range of the KIT. There was no way of determining in advance whether it extended high enough to reach the hypothesized point where dimensionality would begin to rise. Because of these uncertainties, we could not pick out a precise form of relation to predict. If the questionnaire were a single situation and the range of the KIT were too low, we could expect no relation at all to show up. If the questionnaire were a series of situations and the range of the KIT were moderate, we could expect dimensionality to decrease with knowledge. But if the KIT were ample in range, we could expect first a decrease in dimensionality and then an increase. We hoped that our instruments were adequate,

and adopted the latter relation as our Hypothesis 1: Dimensionality in responses to the Problems Questionnaire will first decrease and then increase with score on the KIT.

## RESULTS

### Hypothesis 1

Dividing the KIT scores into six intervals containing as nearly equal numbers of respondents as possible and entering the dimensionality scores for each KIT interval, we obtained the results shown in Table 2; males and females are tabulated separately.

The means shown in Table 2 are plotted graphically in Figure 1. Both

TABLE 2  
FREQUENCIES OF DIMENSIONALITY SCORES  
AT SUCCESSIVE KIT LEVELS

		KIT scores						
		14- 23	24- 26	27- 29	30- 32	33- 36	37- 55	Total
Males only	High 4	2	4	3		2	3	14
	5	5	8	3	4	8	10	38
	6	10	11	5	7	9	6	48
	7	5	5	7	8	2	5	32
	8		3	4	4	3	2	16
	Low 9		2	1	2		4	9
Totals		22	33	23	25	24	30	157
<i>M</i>		5.8	6.0	6.4	6.7	5.8	5.8	
		KIT scores						
		10- 22	23- 26	27- 29	30- 31	32- 34	35- 46	Total
Females only	High 3		1					1
	4	2	5		2	3	1	13
	5	9	5	3	6	5	6	34
	6	3	10	6	7	11	12	49
	7	2	3	8	4	2	4	23
	8	1		2	4	4	5	16
	9			2				2
	Low 10				2	1		3
Totals		18	23	21	25	26	28	141
<i>M</i>		5.3	5.5	6.7	6.4	6.1	6.3	

Note.—Chi square tests from tables collapsed to 3 X 3: Males  $\chi^2 = 11.13$ ,  $p < .026$ ; Females  $\chi^2 = 15.80$ ,  $p < .005$ .

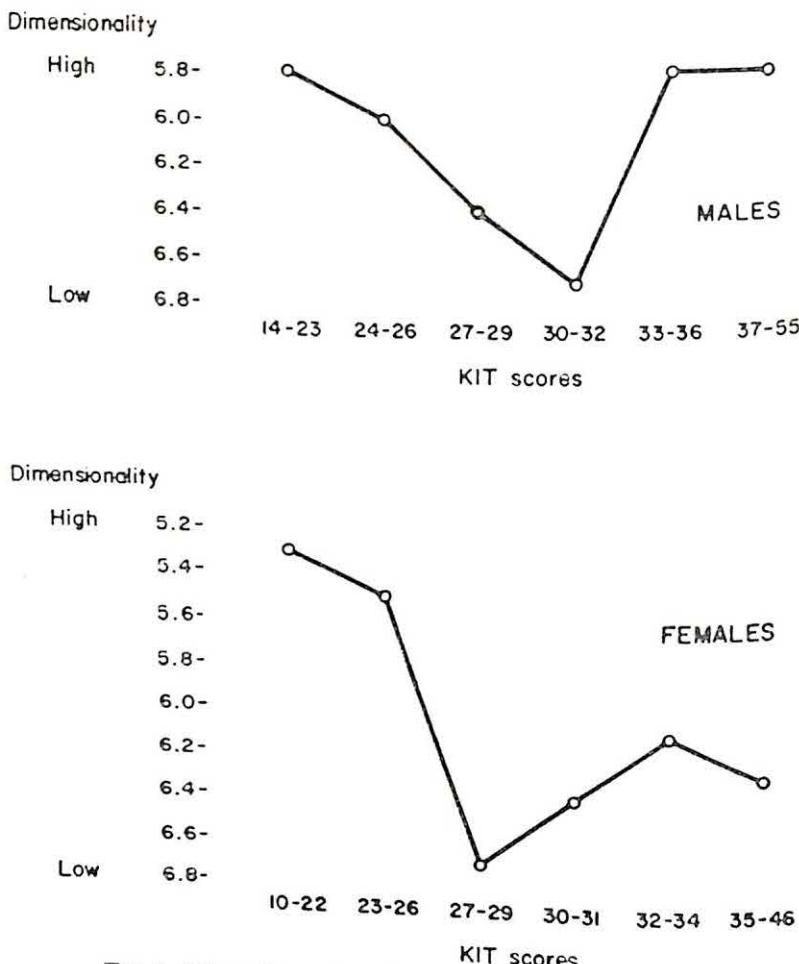


FIG. 1. Mean dimensionality at successive KIT levels.

curves dip in the middle with a gratifyingly high level of statistical confidence, and their general shape is replicated for males and females. Our theory suggests that the high dimensionality of the low KIT scorers represents a haphazard hopping from one dimension to another in going from problem to problem, whereas the moderate to high dimensionality of the high KIT scorers—so let us flatter ourselves—is the differentiated cognitive structure of the expert. The relatively low dimensionality of respondents with average knowledge might be called the simple sophomore world of the semi-educated.

#### Anxiety and Dimensionality

There is evidence in the literature that anxiety focuses or narrows the aspects of the environment to which the organism responds. Easterbrook (1959) is particularly cogent on this point. Now, one way in which perception or attention can be focused is by reducing the number of dimensions along which environmental objects are discriminated; and our dimensionality measure is meant to indicate, roughly, just this: that the types of information presented to the teacher were discriminated along more or fewer dimensions. We should therefore predict (Hypothesis 2) that teachers with higher anx-

iety would show lower dimensionality than those with lower anxiety when responding to the problems.

Our measure of anxiety was the "Self-Perception Checklist" which formed part of a questionnaire used in our larger study. It was made up of selected items from the Guilford-Zimmerman Temperament Survey (1949)<sup>2</sup> and from Saunders' Personality Research Inventory (1955). After the returns were in, these items were analyzed for homogeneity. Thirteen proved to have tetrachoric  $r$ 's with total score of .40 or better, and these were used to make up the total score on anxiety. This score was merely the count of items answered according to the key for "security" as opposed to anxiety, high scores indicating security and the low scores anxiety, the two terms being meant to label a polarity. This short test was admittedly an unreliable instrument and subject to various faults such as social desirability answers; nevertheless, it was the best compromise we could make at the time with costs of administration, respondent fatigue, etc. Some examples of items are these:

Do you frequently get into a state of tension and turmoil by thinking over the day's happenings?

Have you usually been optimistic about your future?

Are you able to criticize people without hurting their feelings?

#### *Hypothesis 2*

When we distributed the security scores against dimensionality with all teachers pooled the result was about as near random as one could get. However, since neither of these measures was expressly designed to give a powerful test of the hypothesis we decided to look further for ways of selecting

respondents for whom the security and dimensionality measures would be more discriminating.

Other work carried out by Gerald Larson (unpublished) as a part of the larger study had suggested that our security measure was more valid for males than for females. Also, as Table 2 shows, the KIT was not only associated with differences in dimensionality, but the spread of dimensionality scores was somewhat greater for the middle group of KIT scores. Accordingly, we distributed security against dimensionality for each of the six groups formed by sex and three levels of KIT. The results for males are shown in Table 3; the results for females were nowhere significant, and are not reported.

For males, results remained random at low and high levels of KIT but became significant in the middle range. In brief, no subgroup of respondents showed a relation contrary to hypothesis; and one subgroup, selected for sensitivity to our unreliable measures, showed the predicted relation at a good level of confidence. As the center subtable of Table 3 shows, relatively more of the secure male teachers exhibited high dimensionality and relatively more of the insecure showed low dimensionality.

#### *Security vs. KIT*

Because both knowledge and anxiety showed some relationship to dimensionality we investigated their relationship with each other. Only random results were obtained for females, but for males the relationship was monotonic. This result is reported in Table 4, where it is shown that more knowledgeable males tend to be more secure than do less knowledgeable ones. Whether the more secure males were those who took courses in testing and learned more from them or whether

<sup>2</sup> Used by permission. We used items from the Guilford-Zimmerman which had been selected in 1954 by Jay M. Jackson.

TABLE 3  
DIMENSIONALITY VS. SECURITY WITHIN  
KIT LEVELS  
Males Only

Dimensionality	Security		Total
	Low (0-10)	High (11-13)	
KIT of 10-25			
High 4-5	10	5	15
6	7	7	14
Low 7-9	5	5	10
Total	22	17	39
$\chi^2$ corrected for continuity = 0.32; ns.			
KIT of 26-35			
High 4-5	7	11	18
6	4	15	19
Low 7-9	20	12	32
Total	31	38	69
$\chi^2$ = 8.54, $p < .01$ (one-tailed). Italicized cell frequencies are higher than randomly expected.			
KIT of 36-54			
High 4-5	4	10	14
6	5	5	10
Low 7-9	2	9	11
Total	11	24	35
$\chi^2$ corrected for continuity = 1.33; ns.			

they became more secure after acquiring more knowledge about testing is, of course, undetermined.

#### SUMMARY

This paper is concerned with the effect of knowledge and of anxiety upon certain aspects of cognitive dimensionality. Dimensionality is here defined as the total number of dimensions or preference factors a person employs in perceiving and evaluating specified complexities of environmental stimuli. Individuals with high di-

mensionality employ a sizable number of dimensions in dealing with such stimuli; those with low dimensionality employ relatively few.

In this research the stimuli consisted of six general kinds of information a teacher might select for use in working with students who presented problems typical of those frequently encountered in the classroom. Our aim was to investigate whether differences in teachers' knowledge about such matters, and also whether differences in teachers' personal feelings of security, affected cognitive dimensionality in ways consistent with those theorized in current literature.

Hypotheses long held by educators combine to assert that training, by serving to increase knowledge, at first restricts and later enlarges the individual's cognitive framework. The untrained person, having no established framework, gives responses primarily determined by the situation. Training first serves to decrease this variability by providing a few specific ways of perceiving these stimuli regardless of the situation, and later serves to enrich cognitive dimensionality by providing the person with knowledge that equips him with more criteria for discriminating the stimuli. The signifi-

TABLE 4  
SECURITY VS. KIT  
Males Only

KIT scores	Security					Total
	0-7	8-9	10-11	12	13	
36-54	2	5	11	9	8	35
26-35	10	15	12	21	11	69
10-25	10	7	10	8	4	39
Total	22	27	33	38	23	143

Note.—Critical ratio by J. E. K. Smith's (1956) monotonicity test is 2.35;  $p < .02$  (two-tailed).

cant curvilinear relationship found between teachers' knowledge of testing and teachers' perceptions of student problems strongly supported this hypothesis. Teachers with low and high levels of knowledge possessed high dimensionality: a finding we interpreted as indicating erratic versus specialized and differentiated perceptions. By contrast, teachers with average knowledge possessed relatively low dimensionality, which we interpreted as indicative of the restricted simplicity of the partially trained individual.

Another body of theory states that feelings of anxiety serve to narrow perception to a few elements of the environment. We reasoned that, if true, such narrowing would be reflected in a reduction of the number of dimensions the anxious individual employs in dealing with his environment. This hypothesis was supported by a part of our data. No relation between anxiety and dimensionality was found for female teachers, but a significant one was found for male teachers who possessed knowledge in the middle or average range. We believe that more reliable measures of anxiety and of dimensionality might provide stronger support for this second hypothesis.

## REFERENCES

- BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, 51, 263-268.
- BIERI, J., & BLACKER, E. The generality of cognitive complexity in the perception of people and inkblots. *J. abnorm. soc. Psychol.*, 1956, 53, 112-117.
- COOMBS, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 471-535.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, 66, 183-201.
- GUILFORD, J. P., & ZIMMERMAN, W. *Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply, 1949.
- LEVENTHAL, H. Cognitive processes and interpersonal predictions. *J. abnorm. soc. Psychol.*, 1957, 55, 176-180.
- LUNDY, R. M., & BERKOWITZ, L. Cognitive complexity and assimilative projection in attitude change. *J. abnorm. soc. Psychol.*, 1957, 55, 34-37.
- SAUNDERS, D. R. *The Personality Research Inventory*. Princeton: Educational Testing Service, 1955.
- SMITH, J. E. K. *On the analysis of contingency tables with ordered classifications*. Lexington, Mass.: Massachusetts Institute of Technology, Lincoln Laboratory, 1956.

(Received March 29, 1961)

# LEARNING IN FOURTH AND SIXTH GRADERS AS A FUNCTION OF SENSORY MODE OF STIM- ULUS PRESENTATION AND OVERT OR COVERT PRACTICE<sup>1</sup>

JOAN LOCKARD

AND

JOSEPH B. SIDOWSKI

*University of Wisconsin*

*San Diego State College*

A number of studies have indicated a relatively consistent trend for subjects to learn more efficiently when material is presented simultaneously to two sensory channels (auditory + visual) than when a single sensory input is used (Day & Beach, 1950). When auditory versus visual modalities are considered, however, results are conflicting. Interpretation of the results is further complicated by the interaction of sensory input variables with other variables. Early studies indicate that auditory presentation of material is more effective than visual for relatively young children (third, fourth, and fifth grades) but that the trend reverses as age increases (Erickson & King, 1917; Russell, 1938). Several recent studies (Evans, Glaser, & Homme, 1960; Silverman & Alter, 1960), investigating response modes in automated instruction with college students, have suggested that the requirement to write a response may be less effective in learning than merely reading the same material. The present study, therefore, was designed in order to investigate the influence of three modes of stimulus presentation (auditory, visual, and auditory + visual) and two modes of responding (overt or covert) upon learning in fourth and sixth grade children.

<sup>1</sup>The authors wish to express their sincere appreciation to Robert Nardelli, Principal of the Campus Laboratory School, San Diego State College, for his cooperation during the course of the experiment.

## METHOD

*Subjects.* Eighteen fourth and 18 sixth grade children from the Campus Laboratory School at San Diego State College served as subjects. The mean IQ (California Tests for Mental Maturity) for the fourth grade was 119.7, and for the sixth grade, 124.6.

*Procedure.* Students in both grades were required to learn lists of nonsense syllables under each of three sensory modes of stimulus presentation: visual, auditory, and visual + auditory. The subjects in each grade level were randomly assigned to two response subgroups, overt or covert, consisting of nine subjects each. In the overt response condition, subjects were required to recall and write down the syllables after each of 14 trials. The covert groups merely learned the syllables and were given a written recall test only after the fourteenth trial. A trial was defined as one complete presentation of the list of 10 syllables.

Five lists of nonsense syllables, two practice and three experimental, taken from Glaze (1928) and of high association value (87-100%) were used. In order to keep intralist similarity at a minimum, a different consonant started each of the syllables in a list and vowels were used about equally often. Each list consisted of 10 syllables. An earlier study by Sidowski and Eason (1960) indicated no significant difference in difficulty between the lists. Each list was presented for a total of 14 trials in each condition. Syllables were presented at a 2-second rate with a 2-second interval intervening between the presentation of each syllable. The syllables were spelled out during the 2 seconds for the auditory conditions by means of a Webcor tape recorder. Visual stimuli were presented at eye level by an American Optical Performance 500 projector at a distance of approximately 8 feet from the seated subject. The intertrial interval was 30 seconds for all conditions. Each subject in the overt groups was required to write down all of the nonsense syllables that he remembered during this interval on each of the 14 trials. Subjects in the covert groups merely sat and waited for the

next trial. All subjects were instructed to remember as many of the syllables as possible since they would be tested after going through the list 14 times. After Trial 14, subjects in all groups were allowed as much time as was needed to recall and write down the syllables.

**Design.** Each subject served in the experiment one hour per day for 4 consecutive days. The first day was a practice day during which the subject learned one practice list auditorily and a second visually. On each of the following 3 days, learning was carried out under one of the three experimental sensory conditions. The presentation of the sensory variables and the three experimental lists were counterbalanced within the design, as were the presentations of the practice conditions.

At the conclusion of the experiment, the subject was required to rank the three sensory variables from the least preferred to the most preferred.

## RESULTS AND DISCUSSION

The mean number of errors in recall are presented in Table 1 for all conditions. (An error was defined as an incorrect recall as well as a failure to recall.) An inspection of the data indicates a consistent superiority for the visual, and visual + auditory presentations for the overt practice condition. This superiority is consistent for both grade levels. The covert practice condition appears superior for the sixth graders with little differentiation at this grade level between the sensory variables.

An analysis of variance was carried out on the data for Trial 14 since recall measures were taken during this trial for the overt and covert conditions. The results of the analysis are presented in Table 2. A significant overall difference was found between the sensory modes of stimulus presentation, while the Grades  $\times$  Response Type interaction showed an  $F$  of 4.11, with an  $F$  of 4.15 being required at the 5% level. Analysis of the sensory mode variable indicated that the auditory presentation of stimuli was sig-

TABLE 1  
MEAN NUMBER OF ERRORS IN RECALL FOR  
OVERT AND COVERT RESPONSE  
CONDITIONS

Stimuli	Response groups		
	Overt		Covert
	Trial 1	Trial 14	Trial 14
Grade 4:			
Auditory	8.2	2.8	3.3
Visual	7.3	0.8	2.1
Auditory + Visual	7.3	2.0	2.3
Grade 6:			
Auditory	8.3	3.2	0.07
Visual	6.6	1.2	0.07
Auditory + Visual	7.1	1.5	0.07

nificantly inferior in recall to visual but not significantly different from auditory + visual.

It was decided beforehand to carry out two additional statistical analyses. A separate analysis was carried out on the covert response groups and the results indicated: (a) fourth graders producing significantly more errors than sixth graders, (b) no significant difference between sensory modes for the sixth grade, and (c) for the fourth grade, auditory presentation producing significantly more errors ( $p < .05$ ) than visual. A final analysis of variance was carried out on Trial 1 for the overt response condition. The results indicated a significant sensory mode effect ( $F = 7.95$ ,  $df = 2/32$ ,  $p < .01$ ) with auditory presentation showing significantly more errors than either of the other sensory modes for both grade levels.

It is obvious that the results of the present study do not agree with those of earlier studies (Erickson & King, 1917; Russell, 1938) which indicate a superiority for the auditory method for younger children. In the earlier studies, however: (a) more meaningful material was used, (b) the audi-

TABLE 2  
ANALYSIS OF VARIANCE FOR ERRORS IN  
RECALL ON TRIAL 14

Source	df	MS	F
Grades (G)	1	22.23	2.83
Response types (RT)	1	2.09	
G × RT	1	32.23	4.11
Subjects within groups	32	7.85	
Sensory modes (SM)	2	15.87	9.12*
SM × G	2	0.84	
SM × RT	2	5.08	2.92
SM × RT × G	2	2.07	1.19
Error	64	1.74	
Total	107		

\* p = .01.

tory presentation of material consisted of class lecture and/or reading by a status figure, and (c) less stringent controls were exercised over the experimental environment. One of the advantages inherent in the use of the visual method for presenting nonsense syllables is that the letters for any three letter syllable are presented simultaneously. With the auditory method, each of the letters is spelled serially thus requiring the subject to organize the letters in the syllable while learning (Krawiec, 1946). Evidently the covert condition allowed the sixth graders to learn and organize the material better than the overt method for all modes of stimulus presentation. This was not true for the fourth graders, although the younger children had little difficulty in recalling the material as well as the sixth graders when written responses were required after each trial.

The results of the preference rankings of the three sensory modes paralleled those of the error score analyses. The visual and auditory + visual modes were preferred significantly more than the auditory mode, with little preferential difference being in-

dicated between the two visual conditions.

### SUMMARY

The learning of nonsense syllables by fourth and sixth graders was investigated as a function of: (a) three modes of stimulus presentation (auditory, visual, and auditory + visual); and (b) two response types, overt (written) and covert (nonwritten).

All subjects were required to learn a list of nonsense syllables under each sensory mode of stimulus presentation. Each list was presented for 14 trials under all conditions. Half of the subjects in each grade recalled and wrote out as many of the syllables as possible after each of the 14 trials (overt group). The remaining subjects merely observed and learned the syllables for 14 trials with no written recall responses (covert). After the 14 trials, subjects in all groups were given as much time as was needed to recall and write out the syllables.

Analyses of the data indicated a significant superiority for the visual and auditory + visual methods of stimulus presentation. No significant difference was found between the fourth and sixth graders for the overt response condition. For the covert groups, however, the fourth graders recalled significantly fewer syllables than the sixth graders.

### REFERENCES

- DAY, W. F., & BEACH, B. R. A survey of the research literature comparing the visual and auditory presentation of information. *USAF tech. Rep.*, 1950, No. 5921. Wright-Patterson Air Force Base, (Dayton, Ohio).
- ELLIOT, F. R. Memory for visual, auditory and visual-auditory material. *Arch. Psychol.*, NY, 1936, 29, No. 199.
- ERICKSON, C. I., & KING, I. A comparison of visual and oral presentation of lessons in the case of pupils from the third to the ninth grade. *Sch. Soc.*, 1917, 6, 146-148.

- EVANS, J. L., GLASER, R., & HOMME, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumdsaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book*. Washington, D. C.: National Education Association, 1960. Pp. 446-451.
- GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255-267.
- KRAWIEC, T. S. A comparison of learning and retention of materials presented visually and auditorily. *J. gen. Psychol.*, 1946, **34**, 179-195.
- LUMLEY, F. H. Research in radio education at Ohio State University. In, *Education on the air: Fourth yearbook of the institute for education by radio*. Columbus: Ohio State University, 1933.
- RUSSELL, R. D. A comparison of two methods of learning. *J. educ. Res.*, 1938, **18**, 235-239.
- SIDOWSKI, J. B., & EASON, R. G. Drive, verbal performance, and muscle action potential. *J. exp. Psychol.*, 1960, **60**, 365-370.
- SILVERMAN, R. E., & ALTER, M. Note on the response in teaching machine programs. *Psychol. Rep.*, 1960, **7**, 496.

(Received April 5, 1960)

## THE ROLE OF DISCRIMINABILITY IN MEANINGFUL VERBAL LEARNING AND RETENTION

DAVID P. AUSUBEL AND DONALD FITZGERALD

*Bureau of Educational Research, University of Illinois*

In a recent study (Ausubel, 1960), it was shown that introductory material at a high level of abstraction, generality, and inclusiveness (advance organizers) facilitates meaningful verbal learning and retention. By deliberately introducing relevant and appropriately inclusive subsuming concepts into cognitive structure, one provides helpful ideational scaffolding which enhances the incorporability and longevity of the more detailed material in the learning passage.

Advance organizers, however, ordinarily have two distinct functions that correspond in turn to two different aspects of the unfamiliarity of meaningful learning material. Sometimes, as in the above mentioned experiment, the new material is almost completely unfamiliar in the sense that cognitive structure is barren of even generally related concepts. Under these circumstances the purpose of the organizer is simply to provide ideational anchorage or scaffolding. More typically, however, the new learning material (e.g., Buddhist doctrines) is a variant of related, previously learned concepts (Christian doctrines) already established in cognitive structure. Here the role of the organizer is not only to provide optimal anchorage at an optimal level of inclusiveness, but also to increase the discriminability of the learning passage from analogous and often conflicting ideas in the learner's cognitive structure.

This second role of organizers is predicated on the assumption that if the distinguishing features of the new learning passage are not originally sa-

lient or readily discriminable from established ideas in cognitive structure, they can be adequately represented by the latter for memorial purposes, and hence would not persist as separately identifiable memories in their own right. It is assumed, in other words, that only discriminable categorical variants of previously learned concepts have long-term retention potentialities. Thus, if a *comparative* type of organizer could first delineate clearly, precisely, and explicitly the principal similarities and differences between the new learning passage (Buddhism) and existing, related concepts in cognitive structure (Christianity), it seems reasonable to suppose that the more detailed Buddhist ideas would be grasped later with fewer ambiguities, fewer competing meanings, and fewer misconceptions suggested by the learner's knowledge of Christianity; and that as these clearer, less confused Buddhist meanings interact with analogous Christianity meanings during the retention interval, they would be more likely to retain their identity. In this experiment, therefore, the value of a comparative organizer was tested by contrasting its effects on the retention of a Buddhism learning passage with both those of a nonideational (*historical*) introduction and those of a simple *expository* organizer.

It is hypothesized, first, that to the extent that the organizer is rendered discriminable from related concepts (Christianity) established in cognitive structure, and hence to the extent that it increases the discriminability of the

Buddhism learning passage from these Christianity concepts, it facilitates the learning and retention of the new Buddhist ideas. In addition, it is hypothesized, for analogous reasons, that the discriminability (and hence the learning and retention) of the Buddhism passage varies as a function of the clarity and stability of the learner's existing knowledge of Christianity, and that subjects with relatively unclear and unstable concepts of Christianity derive relatively more benefit from the organizers than do subjects with clear and stable concepts in this area of knowledge.

These hypotheses were suggested by the findings of a previous experiment (Ausubel & Blake, 1958) in which the learning and retention of a comparative Buddhism passage were contrasted with that of an expository Buddhism passage. However, for the following reasons, the organizer approach was adopted in this experiment, in preference to manipulating the learning material itself: First, organizers provide advance ideational scaffolding. Second, they provide the learner with a generalized overview of all of the major similarities and differences between the two bodies of ideas before he encounters the new concepts individually in more detailed and particularized form. Finally, they create an advance set in the learner to perceive similarities and differences, and, by avoiding overly explicit specification, encourage him actively to make his own differentiations in terms of his own particular sources of confusion.

The hypotheses advanced in this study apply only to meaningful learning material which, although unfamiliar to subjects, is relatable to familiar to subjects, is relatable to long-established and relatively stable concepts in cognitive structure. They do not apply to rote learning, to com-

pletely unfamiliar learning material, or to learning tasks that can only be related to unstable or recently learned concepts. The organizers, furthermore, must consist of ideational material (both similarities and differences) at a high level of abstraction, generality, and inclusiveness, rather than constitute a simple summary or a mere listing of specific comparative points.

## METHOD

### Subjects

The experimental population consisted of predominantly senior undergraduate students (94 women and 61 men) in six sections of an educational psychology course at the University of Illinois. All subjects were enrolled in one of 10 teacher education curricula at the secondary school level. The experiment was conducted separately in each section as a required laboratory exercise, and was performed during regularly scheduled class hours. In order to maximize ego involvement, subjects were informed that after the data were processed their individual scores, as well as the class results, would be reported to them.

### *Learning Passages, Organizers, and Measuring Instruments*

The learning material used in this study was a specially prepared 2,500-word passage dealing with Buddhist concepts of God, immortality, soul, faith, salvation, morality, and responsibility. These concepts were elaborated in considerable detail. The passage only presented the significant *ideas* of Buddhism and ignored material on the life of Buddha, the history and geographical distribution of Buddhism, schools of Buddhism, and Buddhist ritual. Flesch analysis of the passage yielded an abstractness score of 19.19 (highly abstract) and a readability score of 42.46 (difficult).

The topic of Buddhism was chosen both because it was explicitly unfamiliar to undergraduate students, and because it dealt with variants of previously learned concepts (i.e., Christian doctrines) generally familiar to all of our subjects and, presumably, reasonably well-established in the cognitive structure of most of them. Both criteria were important since the main purpose of the study was to

ascertain whether advance organizers could facilitate the learning and retention of *unfamiliar* meaningful material by increasing the discriminability between the new material and *related* concepts already established in cognitive structure. The use of unfamiliar learning material also made it possible for all subjects to start from approximately the same base line in learning the passage. Empirical confirmation of the unfamiliarity of the Buddhism material was obtained when a group of comparable subjects, who had not studied the learning passage, made scores which, on the average, were only slightly and not significantly better than chance.

Still another advantage of the Buddhism material inhered in the fact that its learnability was relatively uninfluenced in our population by such factors as sex, field of specialization, and variability in relevant incidental experience. It is true that women subjects made significantly higher scores than men subjects on both the Buddhism and Christianity tests. This superiority, however, was not a function of differential background, interest, or motivation related to sex per se, but reflected significantly superior academic aptitude as measured by verbal score on the School and College Ability Test (SCAT).

Knowledge of the Buddhism material was tested 3 and 10 days after learning, by equivalent forms of a 45-item multiple-choice test with corrected split-half reliabilities of .80 and .79, respectively. The correlation of the 3- and 10-day retention scores was .79. This latter figure indicated both relatively high stability over time, in view of the different (3- versus 10-day) retention abilities involved, as well as a high degree of relationship between the two retention abilities. Test questions covered principles, facts, and applications, and were selected by an item analysis procedure from a larger population of items. Scores on both tests showed a satisfactory range of variability and were distributed normally. Since they were intended as power tests, no time limit was imposed.

Three types of introductory passages were used in this experiment, each about 500 words in length. The *comparative* organizer pointed out explicitly the principal similarities and differences between Buddhist and Christian doctrines. This comparison was presented at a much higher level of abstraction, generality, and inclusiveness than the Buddhism passage itself, and was deliberately designed to increase discriminability between the two sets of concepts. The *expository* organizer, on the other hand, merely presented the

principal Buddhist doctrines at a high level of abstraction, generality, and inclusiveness, without making any reference whatsoever to Christianity. It was not explicitly designed to increase discriminability between the two religions, but merely to provide some general ideational scaffolding for the detailed Buddhist material. The *historical* introduction was intended solely as a control treatment. It contained interesting historical and human interest material about Buddha and Buddhism, but neither provided any ideational scaffolding nor attempted to compare Buddhism and Christianity. No information was included in any of the introductory passages that could constitute a direct advantage in answering questions on the Buddhism test.

It was methodologically important to provide an historical introduction for the control group in order that any obtained differences in retention outcomes between experimental (comparative) and control (historical) groups could be attributed to the particular nature of the comparative organizer (i.e., to its enhancing effects on discriminability) rather than to its presence per se. The purpose of exposing another group of subjects to an expository introduction was to determine whether the comparative organizer could increase the discriminability of the learning passage over and above that which could be attributed to the influence of a simple organizer (i.e., to the mere provision of advance ideational scaffolding).

A 36-item multiple-choice test on Christianity was used to measure variability among our subjects with respect to the stability and clarity of those existing (Christianity) concepts within cognitive structure which were analogous to the Buddhism learning material, and hence potentially interfering. Test items were noncontroversial, dealing with Old and New Testaments, church history, and denominational beliefs and differences. The scores on this test were normally distributed and the corrected split-half reliability was .84.

Scores on the verbal portion of the SCAT were available for 65 subjects. This test had been previously administered as part of a battery of entrance examinations for incoming university freshmen.

#### Procedure

On the first day of experimentation, all subjects took the Christianity test and then (after assignment to a treatment group) studied one of the three kinds of introductions for 8 minutes. Membership in a treatment group (comparative, expository, or

historical) was determined by random assignment. The population of each of the three treatment groups was also stratified by sex so that the proportion of men to women subjects would be the same in each group. This was necessary because of the women's significantly higher verbal SCAT scores. It was possible to administer all three treatments simultaneously because they consisted of identical appearing introductory passages (with identical sets of directions), differing only in content.

To equalize the possible effects of prior extended exposure to the Buddhism material, those few subjects (5% of the total) who had taken a course in comparative religion were equally distributed in random fashion, among the three treatment groups. It had also been assumed that random assignment of subjects would render the different treatment groups equivalent with respect to such factors as learning ability and knowledge of Christianity. This assumption was confirmed empirically by the finding that differences between these groups on SCAT and Christianity test scores were negligible.

In order to control for the effects of different instructor, situational, and classroom climate variables in the six sections, students within each section were equally divided among the three treatment groups. Since analysis of the data showed that homogeneity of variance prevailed, both on an intersectional as well as on an intergroup basis, for both the Christianity test scores and the two sets of Buddhism test scores, it was considered justifiable to treat the scores of the three treatment groups on each of these instruments as comparable random samples from the same population.

Two days after studying their designated introductions, all subjects read and studied the same Buddhism passage for 35 minutes. One form of the Buddhism test was administered to all subjects 3 days later, and an equivalent form of the same test was administered one week after the first test, or 10 days after the learning passage.

A special randomly assigned control group of subjects was constituted out of the six sections, along the same lines described above for the three treatment groups. The procedure followed with this special group was identical with that used for the comparative group except that subjects studied a 1,800-word passage on the endocrinology of human pubescence instead of the Buddhism passage. The purpose of this procedure was to ascertain to what extent mere knowledge of the comparative organizer (without any expo-

sure to the Buddhism learning passage itself) could increase scores on the Buddhism tests beyond chance expectancy.

## RESULTS AND DISCUSSION

Comparison of corresponding 3- and 10-day means of total treatment groups shows that retention loss during this interval was relatively slight. The loss was greatest in the historical group ( $p = .05$ ), least in the expository group, and intermediate in the comparative group ( $p < .10$ ). This low degree of retention loss is probably attributable both to the negatively accelerated shape of most retention curves, particularly in the case of meaningful material, and to the "rehearsal effect" of the 3-day Buddhism test on the subsequent test of retention. The retention scores of our subjects over the two intervals were highly correlated ( $r = .79$ ).

### *Effect of Organizers on Retention*

On a 3-day basis, only the comparative organizer was effective in facilitating retention of the Buddhism material. Table 1 shows that the mean retention score of the total comparative group was significantly superior to both that of the historical group ( $p = .05$ ) and that of the expository group ( $p = .05$ ). However, practically all of this obtained difference between the comparative group, on the one hand, and the expository and historical groups, on the other, was derived from the below-median subgroups on the Christianity test. Within these below-median subgroups the differences between the means in question were much greater than the corresponding differences between total treatment groups, and their level of significance was also correspondingly higher ( $p = .02$ ,  $p = .005$ ).

It is apparent, therefore, that al-

though provision of ideational scaffolding in the form of an expository organizer did not enhance retention of the Buddhism passage over a 3-day interval, the combined scaffolding and explicit discriminability effects induced by the comparative organizer did significantly improve retention. The short-term retention loss in the control group was evidently small enough to preclude the possibility of a significant difference in retention attributable to the facilitating effects of a simple expository organizer. It is true that an expository type of organizer significantly increased 3-day retention in an earlier experiment (Ausubel, 1960); but the learning passage used then was more unfamiliar to subjects, and the latter also had the benefit of studying the organizer on two separate occasions.

As hypothesized, subjects with relatively superior knowledge of Christianity derived considerably less benefit from the comparative organizer than did subjects whose knowledge of Christianity was less impressive. This finding was consistent with the self-evident proposition that if discriminability of a learning passage is already high because of endogenous factors within cognitive structure (i.e., because of the clarity and stability of related established knowledge), less scope exists for the potentially facilitating influence of exogenously manipulated factors (i.e., organizers) designed to promote discriminability.

On a 10-day basis, both the comparative and expository total groups were significantly superior to the historical total group in retaining the Buddhism material ( $p = .02$ ,  $p = .05$ ), but the difference between the comparative and expository groups was negligible (see Table 1). As was true of the 3-day scores, however, most of the difference between or-

ganizer and control groups was derived from the below-median subgroups on the Christianity test. When the means of just these below-median subgroups were compared, the significance level of the comparative-historical and the expository-historical differences was enhanced ( $p = .0025$ ,  $p < .05$ ), and the difference between the comparative and expository groups was significant at the .10 level.

In comparing the 10- to the 3-day retention data, it appears first, that only over the longer time interval was the natural retention loss sufficiently great to provide scope for the limited facilitating influence of the scaffolding effects available from the expository organizer. Second, although the comparative organizer was not significantly more effective than the expository organizer over the longer interval when the results of all subjects in these groups were considered, there was a suggestive trend in this direction among the below-median subjects.

One explanation of the relatively small difference in 10-day retention attributable to the influence of explicit comparison per se, is the possibility that confronting the learner in advance with the major principles of Buddhism in a detail-free context (the expository organizer) *implicitly* increased the discriminability of his Buddhism ideas by enabling him to make his own comparisons with Christianity. Another plausible explanation is that by enhancing retention generally, the rehearsal effect induced by the 3-day Buddhism test had a leveling influence on the relative degree of facilitation that might have been expected from the two kinds of organizers. This interpretation is supported by the erosion of the significant retention difference between the comparative and expository

TABLE 1  
MEAN RETENTION TEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON BUDDHISM PASSAGE

Knowledge of Christianity	Treatment group		
	Comparative organizer	Expository organizer	Control (historical)
<b>3-day retention:</b>			
Above-median	23.50	22.50	23.42
Below-median	20.50	17.32	16.52
Total	21.83	19.91	19.97
<b>10-day retention:</b>			
Above-median	21.79	22.27	20.87
Below-median	19.21	17.02	14.40
Total	20.31	19.65	17.63

Note.—Chance score on the multiple-choice test of 45 items is 9.0.

groups from the third to the tenth day.

The tendency noted above in the 3-day results—for only the below- as opposed to the above-median subgroups on the Christianity test to derive appreciable benefit from the organizers—also appeared in the 10-day data. Thus, in the learning and retention of unfamiliar ideational material that is relatable to established concepts in the learner's cognitive structure, both comparative and expository organizers appear to be effective only in those instances where existing (endogenously determined) discriminability between the two sets of ideas is inadequate as a consequence of the instability or ambiguity of the established concepts.

The special control group which only studied the comparative organizer (without any exposure to the Buddhism passage itself) made a mean score of 13.20 on the 3-day Buddhism test and a mean score of 13.45 on the 10-day test. These scores were significantly greater than the scores of a comparable naive group (which took the Buddhism tests without being exposed to either organizer

or learning passage), but were substantially below those of the historical and two organizer groups.

It is quite unlikely, however, that the organizers per se directly furnished pertinent information enabling subjects to obtain higher scores on the Buddhism tests. In the first place, a deliberate effort was made to avoid providing such information in the two organizers. Second, if the organizers themselves had furnished useful information in answering test items, it would be difficult to explain why subjects in the expository group did not make higher scores than subjects in the historical group on the 3-day Buddhism test. Much more credible, therefore, is the explanation that exposure to the comparative organizer merely increased the general sophistication of subjects in the special control group about Buddhist concepts, and thereby helped them to eliminate misleads in the multiple-choice test questions. Thus, they were able to obtain better than chance scores without studying the learning passage itself. But when the learning passage was available (i.e., in the historical and two organizer groups), subjects were neither benefited by the general sophistication they obtained from an ideational organizer, nor handicapped by not possessing such sophistication.

#### *Effect of Knowledge of Christianity on Retention*

The data clearly support the hypothesis that the discriminability of the Buddhism learning material varies as a function of the clarity and stability of the established concepts to which it is related (i.e., Christianity), and hence that Buddhism retention scores are positively correlated with knowledge of Christianity. Table 1 shows that within each treatment

TABLE 2  
CORRELATIONS OF BUDDHISM RETENTION  
SCORES AND OTHER MEASURES

Other measures	Buddhism retention scores					
	Comparative group		Expository group		Historical group	
	3-day	10-day	3-day	10-day	3-day	10-day
Christianity scores	.37*	.21	.57**	.42**	.55**	.56**
Verbal ability scores	.62**	.60**	.75**	.79**	.58**	.52*
Christianity scores with verbal ability eliminated	.06	-.07	.40**	.14	.40**	.43**

\* Significant at .05 level.

\*\* Significant at .01 level.

group the mean retention score of the above-median group was significantly greater than the mean retention score of the below-median group. For the historical, expository, and comparative groups, these differences were significant at the .0005, .001, and .02 levels, respectively, on the 3-day Buddhism test, and at the .0005, .0005, and .10 levels, respectively, on the 10-day Buddhism test. The same trends are shown by the positive correlations between scores on the Buddhism and Christianity tests (see Table 2).

It is evident, therefore, that the clearer and more stable subjects' knowledge of Christianity was, the more discriminable this knowledge was from Buddhism concepts, and hence the higher the Buddhism retention scores were. And conversely, the less clear and more unstable subjects' knowledge of Christianity was, the less discriminable it was from Buddhism concepts, and the lower the Buddhism retention scores were.

The data also confirm the hypothesis that organizers (by virtue of their

leveling effect on the endogenous discriminability advantage inherent in a clear and stable knowledge of Christianity) reduce the relationship between Christianity knowledge and Buddhism retention scores in proportion to their facilitating effect on retention. The leveling effect of the organizers (particularly the comparative) can be seen in Table 1 by comparing the relative magnitude of the differences between above-median and below-median subgroups in the various treatment groups. For historical, expository, and comparative groups the respective *t* values of these differences were 4.86, 3.43, and 2.24 for the 3-day Buddhism scores, and 4.72, 3.99, and 1.59 for the 10-day Buddhism scores. Correlations between the Buddhism and Christianity scores show the same trend (see Table 2). The correlation between the Christianity and Buddhism scores was higher in the historical than in the comparative group for both the 3-day ( $p = .12$ ) and 10-day ( $p = .04$ ) retention intervals. The almost identical correlations, on the other hand, between Christianity and 3-day Buddhism scores in the expository and historical groups, reflected the ineffectiveness of the expository organizer in facilitating retention over the 3-day interval.

Because of the positive correlation of .44 ( $p < .01$ ) between Christianity and SCAT scores, however, it was necessary to check the alternative hypothesis that the significantly higher retention scores of subjects in the above-median groups reflect superior verbal ability rather than superior knowledge of Christianity per se (endogenous discriminability). But since the availability of only 65 SCAT scores made it impossible to match for verbal ability those subjects within each treatment group who were above

and below the median score on the Christianity test, it was necessary to resort to partial correlation.

The partial correlations between the Buddhism and Christianity scores, with the effect of verbal ability eliminated, are shown for all three groups in Table 2. For the historical group the partial correlations of .40 and .43 (for the 3- and 10-day scores, respectively), although lower than the corresponding simple correlations of .55 and .56 were still significant at the .01 level. Three of the four other partial correlations, however, were negligible and nonsignificant, indicating that the corresponding simple correlations largely reflected the positive relationship between verbal ability and Buddhism retention scores. The intrinsic residual relationship between knowledge of Christianity and retention of the Buddhism material, after the effect of verbal ability was eliminated, was actually close to zero in these instances, because of the leveling influence of the organizers on the endogenous discriminability advantage conferred by superior knowledge of Christianity. The partial correlation between the Christianity and 3-day Buddhism scores remained significant at the .01 level in the expository group, in as much as the expository organizer did not facilitate retention over the 3-day interval.

Intergroup comparison of the partial correlations in Table 2 shows even more clearly than does corresponding comparison of the simple correlations between Buddhism and Christianity scores, that the organizers (in proportion to their facilitating effect on retention) reduced the significant relationship between the significant relationship between knowledge of Christianity and retention of the Buddhism material. With the influence of verbal ability eliminated, the partial correlation between

Christianity and 3-day Buddhism scores in both the historical and expository groups was significantly higher than in the comparative group ( $p = .05$ ,  $p = .05$ ). The partial correlation between Christianity and 10-day Buddhism scores in the historical group was higher than the corresponding partial correlations in the expository ( $p = .08$ ) and comparative ( $p = .01$ ) groups. In effect then, by being provided with a given type of discriminability aid (organizer), subjects possessing relatively little knowledge of Christianity (and hence little endogenous discriminability) were placed on approximately the same footing with respect to the discriminability variable and its effect on retention, as subjects possessing greater knowledge of Christianity (and greater endogenous discriminability).

Table 2 also shows that the correlation between verbal ability and Buddhism retention scores in the historical group (where the relationship was not influenced by interaction with organizer effects) was approximately the same as that between Christianity and Buddhism scores. But unlike the latter relationship, which was attenuated by interaction with the organizers, the correlation between verbal ability and Buddhism scores was slightly higher in the comparative than in the historical group, and suggestively but not significantly higher in the expository than in the historical group ( $p = .17$ ,  $p = .07$ ). At the very least, therefore, it is definite that neither organizer detracted from the strong positive relationship between verbal ability and the retention of the Buddhism material; and there was a suggestive tendency for verbal ability to have an even greater impact on the retention of the Buddhism passage

when subjects were given the benefit of an expository organizer.

#### SUMMARY AND CONCLUSIONS

This experiment was concerned with the role of endogenous and externally manipulated discriminability in the learning and retention of unfamiliar ideational material (Buddhism) that was relatable to previously learned concepts (Christianity) already established in cognitive structure.

The learning task consisted of a 2,500-word passage on the principles of Buddhism. Two days before studying this learning passage, one experimental group studied a 500-word *comparative* organizer explicitly comparing the major ideas of Buddhism and Christianity. Another experimental group studied an *expository* organizer which made no reference to Christianity. A control group studied an *historical* introduction dealing with the history rather than with the ideas of Buddhism. Retention of the Buddhism material was tested 3 and 10 days after the learning session by means of equivalent forms of a multiple-choice test. For purposes of analysis, subjects in the various treatment groups (university undergraduates) were divided into above- and below-median subgroups in terms of their scores on an objective test of Christianity.

Subjects with greater knowledge of Christianity made significantly higher scores on the Buddhism retention test than did subjects with less knowledge of Christianity. This significantly positive relationship be-

tween Christianity and Buddhism test scores held up even when the effect of verbal ability was eliminated. Hence the data support the hypothesis that the learning and retention of unfamiliar verbal material varies positively with its discriminability from related, previous learned concepts established in cognitive structure, and that this endogenously determined discriminability is a function of the clarity and stability of the latter concepts.

On a 3-day basis only the comparative organizer was significantly effective in facilitating the retention of the Buddhism material, but over a 10-day interval both comparative and expository organizers were significantly effective. These facilitating effects of the organizers on retention outcomes, however, only applied to subjects in the below-median subgroups on the Christianity test. Thus in the learning and retention of unfamiliar ideational material that is relatable to established concepts in the learner's cognitive structure, both comparative and expository organizers appear to be effective only in those instances where existing discriminability between the two sets of ideas is inadequate as a consequence of the instability or ambiguity of established concepts.

#### REFERENCES

- AUSUBEL, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. *J. educ. Psychol.*, 1960, **51**, 267-272.  
AUSUBEL, D. P., & BLAKE, E., JR. Proactive inhibition in the forgetting of meaningful school material. *J. educ. Res.*, 1958, **52**, 145-149.

(Received May 8, 1961)

# JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 52

December 1961

Number 6

## THE GENETIC APPROACH TO THE PSYCHOLOGY OF THOUGHT<sup>1</sup>

JEAN PIAGET

*University of Geneva, Switzerland*

From a developmental point of view, the essential in the act of thinking is not contemplation—that is to say, that which the Greeks called "theorema"—but the action of the dynamics.

Taking into consideration all that is known, one can distinguish two principal aspects:

1. The formal viewpoint which deals with the configuration of the state of things to know—for instance, most perceptions, mental images, imageries.

2. The *dynamic* aspect, which deals with transformations—for instance, to disconnect a motor in order to understand its functioning, to disassociate and vary the components of a physical phenomenon, to understand its causalities, to isolate the elements of a geometrical figure in order to investigate its properties, etc.

The study of the development of thought shows that the dynamic aspect is at the same time more difficult to attain and more important, because only transformations make us under-

stand the state of things. For instance: when a child of 4 to 6 years transfers a liquid from a large and low glass into a narrow and higher glass, he believes in general that the quantity of the liquid has increased, because he is limited to comparing the initial state (low level) to the final state (high level) without concerning himself with the transformation. Toward 7 or 8 years of age, on the other hand, a child discovers the preservation of the liquid, because he will think in terms of transformation. He will say that nothing has been taken away and nothing added, and, if the level of the liquid rises, this is due to a loss of width, etc.

The formal aspect of thought makes way, therefore, more and more in the course of the development to its dynamic aspect, until such time when only transformation gives an understanding of things. To think means, above all, to understand; and to understand means to arrive at the transformations, which furnish the reason for the state of things. All development of thought is resumed in the following manner: a construction of operations which stem from actions and a gradual subordination of formal aspects into dynamic aspects.

The operation, properly speaking, which constitutes the terminal point of this evolution is, therefore, to be

<sup>1</sup>This paper, transmitted by Ernest Harms, is the text of an address made by Jean Piaget, from notes, at the New York Academy of Sciences' Conference on the Psychology of Thinking on April 28-29, 1960. The translation was made by Ruth Golbin, who has skillfully succeeded in retaining Piaget's individualistic style of expression.

conceived as an internalized action reversible (example: addition and subtraction, etc.) bound to other operations, which form with it a structured whole and which is characterized by well defined laws of totality (example: the groups, the lattice, etc.). Dynamic totalities are clearly different from the "gestalt" because those are characterized by their non-additive composition, consequently irreversible.

So defined, the dynamics intervene in the construction of all thought processes; in the structure of forms and classifications, of relations and serialization of correspondences, of numbers, of space and time, of the causality, etc. One could think at first glance that space and geometry add to the formal aspect of thought. In this way one conceived of the geometric science in the past, considering it impure mathematics, but applicable to perception and intuition. Modern geometry, since *Le Programme d'Erlangen* by F. Klein, has tended, like all other precise disciplines, to subordinate the formal to the dynamic. The geometries are, indeed, understood today as relying all on groups of transformation, so that one can go from one to the other by characterizing one less general "subgroup" as part of a more inclusive group. Thus geometry too rests on a system of dynamics.

Any action of thought consists of combining thought operations and integrating the objects to be understood into systems of dynamic transformation. The psychological criteria of this is the appearance of the notion of conservation or "invariants of groups." Before speech, at the purely sensory-motor stage of a child from 0 to 18 months, it is possible to observe actions which show evidence of such tendencies. For instance: From 4-5 to

18 months, the baby constructs his first invariant, which is the schema of the permanent object (to recover an object which escaped from the field of perception). He succeeds in this by coordinating the positions and the displacements according to a structure, which can be compared to what the geometers call "group displacements."

When, with the beginning of the symbolic function (language, symbolic play, imagerie, etc.), the representation through thought becomes possible, it is at first a question of reconstructing in thought what the action is already able to realize. The actions actually do not become transformed immediately into operations, and one has to wait until about 7 to 8 years for the child to reach a functioning level. During this preoperative period the child, therefore, only arrives at incomplete structures characterized by a lack in the notion of combinations and, consequently, by a lack of logic (in transitivity, etc.).

In the realm of causality one can especially observe these diverse forms of precausality, which we have previously described in detail. It is true that a certain number of authors—Anglo-Saxon above all—have severely criticized these conclusions, while others have recognized the same facts as we have (animism, etc.). Yet, in an important recent book (which will appear soon) two Canadian authors, M. Laurendeau and A. Pinard, have taken the whole problem up once again by means of thorough statistics. In the main points they have come to a remarkable verification of our views, explaining, moreover, the methodological reasons for the divergencies among the preceding authors.

At about 7 to 8 years the child arrives at his first complete dynamic structures (classes, relations, and

numbers), which, however, still remain concrete—in other words, only at the time of a handling of objects (material manipulation or, when possible, directly imagined). It is not before the age of 11 to 12 years or more that operations can be applied to pure hypotheses. At this latter level, a logic of propositions helps complete the concrete structures. This enlarges the structures considerably until their disposition.

The fundamental genetic problem of the psychology of thought is hence to explain the formation of these dynamic structures.

Practically, one would have to rely on three principal factors in order to explain the facts of development: maturation, physical experience, and social interaction. But in this particular case none of these three suffice to furnish us with the desired explanations—not even the three together.

*Maturation.* First of all, none of these dynamic structures are innate, but they form very gradually. (For example: The transitivity of equalities is acquired at approximately 6½ to 7 years, and the ability of linear measure comes about only at 9 years, as does the full understanding of weights, etc.) But progressive construction does not seem to depend on maturation, because the achievements hardly correspond to a particular age. Only the order of succession is constant. However, one witnesses innumerable accelerations or retardations for reasons of education (cultural) or acquired experience. Certainly one cannot deny the inevitable role which maturation plays, but it is determined above all by existing possibility (or limitation). They still remain to be actualized, which brings about other factors. In addition, in the domain of thought, the factors of innateness

seem above all limitative. We do not have, for example, an intuition of space in the fourth dimension; nevertheless we can deduce it.

*Physical Experience.* Experiencing of objects plays, naturally, a very important role in the establishment of dynamic structures, because the operations originate from actions and the actions bear upon the object. This role manifests itself right from the beginning of sensory-motor explorations, preceding language, and it affirms itself continually in the course of manipulations and activities which are appropriate to the antecedent stages. Necessary as the role of experience may be, it does not sufficiently describe the construction of the dynamic structures—and this for the following three reasons.

First, there exist ideas which cannot possibly be derived from the child's experience—for instance, when one changes the shape of a small ball of clay. The child will declare, at 7 to 8 years, that the quantity of the matter is conserved. It does so before discovering the conservation of weight (9 to 10 years) and that of volume (10 to 11 years). What is the quantity of a matter independently of its weight and its volume? This is an abstract notion corresponding to the "substance" of the pre-Socratic physicists. This notion is neither possible to be perceived nor measurable. It is, therefore, the product of a dynamic deduction and not part of an experience. (The problem would not be solved either by presenting the quantity in the form of a bar of chocolate to be eaten.)

Secondly, the various investigations into the learning of logical structure, which we were able to make at our International Center of Genetic Epistemology, lead to a very unanimous

result:<sup>2</sup> one does not "learn" a logical structure as one learns to discover any physical law. For instance, it is easy to bring about the learning of the conservation of weight because of its physical character, but it is difficult to obtain the one of the transitivity of the relationship of the weight:

$$A = C \text{ if } A = B \text{ and } B = C$$

or the one of the relationship of inclusion, etc. The reason for this is that in order to arrive at the learning of a logical structure, one has to build on another more elementary logical (or prelogical) structure. And such structures consequently never stem from experience alone, but suppose always a coordinating activity of the subject.

Thirdly, there exist two types of experiences:

1. The physical experiences show the objects as they are, and the knowledge of them leads to the abstraction directly from the object (example: to discover that a more voluminous matter is more or less heavy than a less voluminous matter).

2. The logicomathematical experience supposes to interrelate by action individual facts into the world of objects, but this refers to the result of these actions rather than to the objects themselves. These interrelations are arrived at by process of abstractions from the actions and their coordinates. For instance, to discover that 10 stones in a line always add up to 10, whether they are counted from left to right or from right to left. Because then the order and the total sum have been presented. The new knowledge consists simply in the discovery that the action of adding a sum is independent of the action of putting them in order. Thus the logicomath-

ematical experience does not stem from the same type of learning as that of the physical experience, but rather from an equilibration of the scheme of actions, as we will see.

*Social Interaction.* The educative and social transmission (linguistic, etc.) plays, naturally, an evident role in the formation of dynamic structures, but this factor does not suffice either to entirely explain its development, and this for two reasons:

First, a certain number of structures do not lend themselves to teaching and are prior to all teaching. One can cite, as an example, most concepts of conservation, of which, in general, the pedagogues agree that they are not problematic to the child.

The second, more fundamental, reason is that in order to understand the adult and his language, the child needs means of assimilation which are formed through structures preliminary to the social transmission itself—for instance, an ancient experience has shown us that French-speaking children understand very early the expression "*quelques unes de mes fleurs*" [some of my flowers] in contrast to "*toutes mes fleurs*" [all my flowers], and this occurs when they have not yet constructed the relation of inclusion:

Some A are part of all B; therefore

$$A < B$$

In conclusion, it is not exaggerated to maintain that the basic factors invoked before in order to explain mental development do not suffice to explain the formation of the dynamic structures. Though all three of them certainly play a necessary role, they do not constitute in themselves sufficient reason and one has to add to them a fourth factor, which we shall try to describe now.

<sup>2</sup> See *Etudes d'Epistomologie Genétique*, Vol. 7 and 10.

This fourth factor seems to us to consist of a general progression of equilibration. This factor intervenes, as is to be expected, in the interaction of the preceding factors. Indeed, if the development depends, on one hand, on internal factors (maturation), and on the other hand on external factors (physical or social), it is self-evident that these internal and external factors equilibrate each other. The question is then to know if we are dealing here only with momentary compromises (unstable equilibrium) or if, on the contrary, this equilibrium becomes more and more stable. This shows that all exchange (mental as well as biological) between the organisms and the milieu (physical and social) is composed of two poles: (a) of the *assimilation* of the given external to the previous internal structures, and (b) of the *accommodation* of these structures to the given ones. The equilibrium between the assimilation and the accommodation is proportionately more stable than the assimilative structures which are better differentiated and co-ordinated.

It is this equilibrium between the assimilation and accommodation that seems to explain to us the functioning of the reversible operations. This occurs, for instance, in the realm of notions of conservation where the invariants of groups do not account for the maturation and the physical experience, nor for the sociolinguial transmission. In fact, dynamic reversibility is a compensatory system of which the idea of conservation constitutes precisely the result. The equilibrium (between the assimilation and the accommodation) is to be defined as a compensation of exterior disturbances through activities of the subject orientated in the contrary direction of

these disturbances. This leads us directly to the reversibility.

Notice that we do not conceive of the idea of equilibrium in the same manner as the "gestalt theory" does, which makes great use of this idea too, but in the sense of an automatical physical equilibrium. We believe, on the contrary, that the mental equilibrium and even the biological one presumes an activity of the subject, or of the organism. It consists in a sort of matching, orientated towards compensation—with even some overcompensation—resulting from strategies of precaution. One knows, for instance, that the homeostasis does not always lead to an exact balance. But it often leads to overcompensation, in response to exterior disturbances. Such is the case in nearly all occurrences except precisely in the case of occurrences of a superior order, which are the operations of reversible intelligence, the reversible logic of which is characterized by a complete and exact compensation (inverted operation).

The idea of equilibrium is so close to the one of reversibility that G. Brunner, in a friendly criticism of one of our latest books appearing in the *British Journal of Psychology*, proposes to renounce the idea of equilibrium because the notion of the reversibility seems sufficient to him. We hesitate to accept this suggestion for the following three reasons:

First, reversibility is a logical idea, while the equilibrium is a causal idea which permits the explanation of reforms by means of a probabilistic schema. For instance, in order to explain the formation of the idea of conservation, one can distinguish a certain number of successive stages, of which each is characterized by the "strategy" of a progress of compensa-

tion. Now it is possible to show<sup>3</sup> that the first of these strategies (only bearing upon one dimension, to the neglect of others) is the most probable at the point of departure, and further, that the second of these strategies (with the emphasis on a second dimension) *becomes* the most likely—as a function of the result of the first. And, finally, that the third of these strategies (oscillation between the observed modifications upon the different dimensions and the discovery of their solidarity) *becomes* the most likely in the functioning of the results of the preceding, etc. From such a point of view the process of equilibration is, therefore, characterized by a sequential control with increasing probabilities. It furnishes a beginning for causal explanations of the reversibility and does not duplicate the former idea.

Secondly, the tendency of equilibrium is much broader for the operation than the reversibility as such, which leads us to explain the reversibility through the equilibrium and not the reverse. In effect, it is at this level of the obvious regulations and sensory-motor feedbacks that the process of equilibration starts. This in its higher form becomes intelligence. Logical reversibility is therefore conceivable as an end result and not as a beginning and the entire reversibility follows the laws of a semireversibility of various levels.

Thirdly, the tendency to equilibrate does not only explain this final reversibility, but also certain new synthesis between originally distinct operations. One can cite in this regard an example of great importance: the serial of whole numbers. Russell and Whitehead have tried to explain the basic set of numbers through the idea

of equivalent classes, without recourse to the serial order. This means that two classes are believed to be equivalent, if one can put their respective elements into a reciprocal arrangement. Only when this relationship relies on the quality of the objects (an A put into relation with an A, a B with a B, etc.) one does not get the quantity. If this relationship is made exclusive of the qualities (an Individual A or B put into relationship with an Individual B or A) then there exists only one way to distinguish the elements from each other. In order not to forget one, or not to count the same twice, one must deal with them in succession and introduce the serial factor as well as the structure of classes. We may then say, psychologically speaking, that the sequence of whole numbers is synthesis between two groupings qualitatively distinct, the fitting of the classes and serialization, and that this synthesis takes place as soon as one excludes the qualities of the elements in question. But how does this synthesis occur? Precisely by a gradual process of equilibration.

On the one hand the child who develops his ideas from numbers is in possession of structures enabling him to fit them into classes (classifications). But if he wants to be exclusive of qualities in order to answer to the question "how many," he becomes unable to distinguish the elements. The disequilibrium which appears, therefore, obliges the child to resort to the idea of order and take recourse to arranging these elements into a lineal row. On the other hand, if the child arranges the elements as 1, 1, 1, etc., how would he know, for instance, how to distinguish the second from the third? This new disequilibrium brings him back to the idea of classification: The "second" is the element which

<sup>3</sup> *Logique et Equilibre*, Vol. 2 of *Etudes d'Epistemologie Genetique*.

has but one predecessor, and the "third" is one that has two of them. In short, every new problem provokes a disequilibrium (recognizable through types of dominant errors) the solution of which consists in a re-equilibration, which brings about a new original synthesis of two systems, up to the point of independence.

During the discussion of my theories, Brunner has said that I have called disequilibrium what others describe as motivation. This is perfectly true, but the advantage of this language is to clarify that a cognitive or

dynamic structure is never independent of motivational factors. The motivation in return is always solidary to structural (therefore cognitive) determined level. The language of the equilibrium presents that activity, that permits us to reunite into one and the same totality those two aspects of behavior which always have a functional solidarity because there exists no structure (cognition) without an energizer (motivation) and vice versa.

(Received May 2, 1961)

## NEED ACHIEVEMENT SCALES AS PREDICTORS OF ACADEMIC SUCCESS

MERVILLE C. SHAW

*Chico State College*

The purpose of the present study was to determine whether or not three of the most widely used scales of need achievement would make appropriate differentiations between groups of high school students matched for ability but differing significantly in mean grade point average. The three scales used in the present research were the Need Achievement scale of the Edwards Personal Preference Schedule, the McClelland Achievement Motivation Test (MAMT), and the French (1956) Achievement scale. Past studies of a similar nature which have made use of need achievement scales have not always been in agreement with each other. McClelland, Atkinson, Clark, & Lowell (1953, p. 238) report a significant positive correlation between their test and grade point average while Lowell (1952) reports non-significant correlations between the same two variables.

Another study by Weiss, Wertheimer, and Groesbeck (1960) reported significant correlations between the EPPS Need Achievement scale and the MAMT as well as between each of these scales and the grade point averages of 60 college students in introductory psychology classes. Atkinson (1950) reports a negative but insignificant relationship between MAMT results and grade point average of college students. A study by Morgan (1952) was similar in design to the present study in that two groups of subjects, matched for ability, but differing significantly in grade point averages, were compared on an MAMT. The *t* test yielded a difference between the two groups on the

MAMT which was significant at the .02 level.

### PROCEDURE

Subjects were selected from the junior and senior classes of a high school of 1,800 students. All students who achieved an IQ above 110 on the California Test of Mental Maturity were considered potential subjects for inclusion in the sample. Students who met the ability criterion and who achieved a grade point average in the previous year of 2.00 (B) or better were classified as achievers. Those who achieved a grade point average of 1.5 or less were classified as underachievers. A total of 18 male achievers and 21 female achievers participated. Twenty male underachievers and 19 female underachievers, likewise, cooperated. All subjects were administered the McClelland Achievement Motivation Test, the Edwards Personal Preference Schedule, and the French Test of Achievement Motivation.

The MAMT was administered to groups through use of an opaque projector under neutral conditions. Cards 7 BM and 8 BM from the TAT and Cards D and H of McClelland's series were the picture stimuli. Subjects responded to the four basic questions used by McClelland. Each MAMT response was scored twice by the author according to the criteria described by McClelland et al. (1953, pp. 107-138). The achievement level of subjects was not known to the scorer at the time protocols were evaluated.

Because of the possibility of low inter-scorer reliability on the French test, three judges were used in the scoring procedure. A given scoring was considered accurate only when at least two of the three judges responded in identical ways. Judges were not aware of the achievement level of the subjects they were judging.

For purposes of statistical analysis, male and female groups were considered separately. Variance ratio (*F*) and *t* tests were applied to the scores of achiever and underachiever groups on each of the three need achievement measures used in order to establish the existence of homogeneity of

TABLE 1  
SIGNIFICANCE OF DIFFERENCE BETWEEN  
MALE AND FEMALE ACHIEVERS AND  
UNDERACHIEVERS ON THE  
CALIFORNIA TEST OF  
MENTAL MATURITY

Group	CTMM IQ means		<i>F</i>	<i>t</i>
	Achievers	Under-achievers		
Male	116.1	115.7	1.108	1.756
Female	117.0	116.7	1.796	1.205

variance and to determine the significance of differences between groups on these measures. In addition, Spearman rank-order correlations between each of the need achievement measures were computed for each of the subgroups in the study in order to determine the extent to which the different achievement measures were measuring the same variable. Finally, an analysis of variance was carried out using the MAMT pictures, achievement level, and sex as differentiating criteria.<sup>1</sup>

### RESULTS

The results reported in Table 1 indicate that groups did not differ in performance on the California Test of Mental Maturity. Analysis of the difference between achiever and underachiever groups on each of the three measures of need achievement is reported in Table 2. The only significant difference found was between male achievers and underachievers on the French Need Achievement scale. In none of the other comparisons was the difference between groups found to be significant.

Computation of correlations between each of the three scales for each group used in the study revealed few significant correlations. Table 3 reports these results.

<sup>1</sup> The author would like to express his sincere appreciation to Richard Rankin, Assistant Professor of Psychology, Chico State College, for computing the analysis of variance.

TABLE 2  
SIGNIFICANCE OF DIFFERENCES BETWEEN  
ACHIEVER AND UNDERACHIEVER GROUPS  
IN THE MAMT, N ACH SCALE OF THE  
EPPS, AND FRENCH N ACH SCALE

Test	Mean score on criterion		<i>F</i>	<i>t</i>
	Achievers	Under-achievers		
MAMT				
Male	5.3	3.5	1.50	1.3503
Female	4.86	3.7	1.46	.8028
EPPS Need Achievement				
Male	16.3	14.6	1.78	1.5889
Female	12.3	12.1	1.17	.1923
French Need Achievement				
Male	8.4	5.9	1.41	2.3640*
Female	7.6	7.2	1.62	.5677

\*  $p < .05$ .

TABLE 3  
RANK-ORDER CORRELATIONS BETWEEN  
EACH OF THE NEED ACHIEVEMENT SCALES  
FOR EACH GROUP INCLUDED IN  
THE STUDY

Group	EPPS & MAMT	EPPS & French	MAMT & French
Males			
Achievers	.12	.51*	.25
Underachievers	-.03	.26	.09
Females			
Achievers	.14	-.15	.28
Underachievers	-.23	-.13	.20

\*  $p < .05$ .

The analysis of variance revealed that variation was highly significant (.001) among the four separate pictures of the MAMT but not significant at the .05 level between either achievers and underachievers or between males and females. The same lack of significance was found to exist between all interaction variances as well.

### DISCUSSION

Results obtained from the analysis of differences between achievers and underachievers on the MAMT and the EPPS Need Achievement scales would seem to indicate that neither of these instruments would have had predictive value for the groups included in this study. The French Need Achievement scale did reveal a difference significant at the 5% level for male members of the sample.

The single significant correlation was found between the Need Achievement scale of the EPPS and the French Need Achievement scale for male achievers. These three scales not only tend to be poor predictors of academic achievement for this group but are not measuring the same variable.

Results of the analysis of variance of the MAMT indicate that the greatest source of variance exists between pictures rather than between groups of achievers and underachievers, between sexes, or between interaction variances. It would appear that for the groups used in this study the four pictures of the MAMT were not all measuring the same qualities.

It would not be reasonable to conclude on the basis of this single study that present need achievement scales are not predictive of academic attainment, but the fact that other studies also have frequently failed to find any relationship between need achievement scales and academic performance signifies the need for a critical re-examination, not only of need achievement instruments, but of the concept of need achievement itself. The basic assumption underlying the idea of the "need to achieve" is that it is a quantity which some people have in greater degree than do others, and that quantitative measures of this variable will provide an index to the

individual's need to achieve in all areas. It would appear that certain fallacies grow out of this assumption. The first fallacy is that the need to achieve is constant for all areas of an individual's life space. The second fallacy is that the need to achieve results in observable production of a socially desirable sort. The third fallacy involved in the basic assumption of the need achievement theory is that felt need to achieve and behavior are congruent. The fact that the intervening variables in the life situation of any given individual may enter into the situation to prevent an individual from meeting any need to achieve he might feel is thus overlooked.

Motives have both force and direction. Present measures of need achievement consider only the former while neglecting the latter. If one is willing to define motivation as the amount of energy expended, then the concept embodied in the idea of "need to achieve" is legitimate. If one believes that motivation cannot be measured in terms of energy output nor amount of production, then the concept of the "need to achieve" will have limited usefulness.

### SUMMARY

Subjects matched for ability were classified as achievers or underachievers on the basis of their academic grades. After completing the McClelland Achievement Motivation Test, the French Need Achievement scale, and the Edwards Personal Preference Schedule groups of male and female achievers and underachievers were compared. Results indicated that none of the three need achievement scales differentiated achievers and underachievers significantly, with the exception of the French scale which did differentiate between male achievers and underachievers. The only signifi-

cant correlation was found between the French and the Edward's need achievement scores for male achievers. Analysis of variance indicated that the greatest source of variance existed among the pictures, rather than between groups of achievers and under-achievers, or between sexes. Implications of these findings as they related to the concept of need achievement were presented.

## REFERENCES

- ATKINSON, J. W. Studies in projective measurement of achievement motivation. Unpublished doctoral dissertation, University of Michigan, 1950.
- EDWARDS, A. W. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1954.

FRENCH, ELIZABETH. Development of a measure of complex motivation. *USAF Personnel Train. Res. Cent. res. Rep.*, 1956, No. TN-56-48.

LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31-40.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. (Eds.) *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MORGAN, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, **16**, 292-298.

WEISS, P., WERTHEIMER, M., & GROESBECK, B. Achievement motivation, academic aptitude and college grades. *Educ. psychol. Measmt.*, 1960, **20**, 663-665.

(Received February 20, 1961)

# THE EFFECTS OF DIRECT AND INDIRECT TEACHER INFLUENCE ON DEPENDENT-PRONE STUDENTS LEARNING GEOMETRY<sup>1</sup>

EDMUND AMIDON

*Temple University*

AND

NED A. FLANDERS

*University of Michigan*

Whether or not a particular type of student can learn when he is exposed to a particular style of teaching has interested a number of researchers. Smith (1955) and Wispe (1951) have both shown that when students are classified by the use of personality test data, they respond differently to highly organized versus loosely organized classroom activities in a college remedial reading course (Smith, 1955) and to college lecturing versus group discussion techniques (Wispe, 1951) in freshman sociology. The present project was concerned with dependent-prone eighth grade students who were exposed to consistently direct versus indirect styles of teaching while learning geometry.

Asch (1951), Kagen and Mussen (1956), and Livson and Mussen (1957) have studied the reactions of dependent-prone persons in various kinds of experimental situations. They concluded that dependent-prone individuals are more likely to comply with authority figures and conform to group pressures than the less dependent-prone. Their results suggest that a dependent-prone student might become overly concerned with following the suggestions and directions of a teacher and more dependent on support and encouragement. The present project was designed to find out if these concerns inhibit or enhance the

learning of geometry at the age level represented in the eighth grade.

## PROCEDURE

This study employed a laboratory design in order to exercise experimental control of spontaneous behavior. First, the behavior of the teacher was controlled by training a teacher-role player. His statements were classified by an observer to demonstrate that desired differences were great enough for students to notice. Second, those students who scored high on a dependence-proneness test, developed by Flanders, Anderson, and Amidon (1960) were selected for the experimental population. Third, control of learning was accomplished by using pre- and posttests of geometry achievement. And fourth, in half of the experimental groups the basic content material, presented in a tape recording, was so organized that the immediate learning goals were unclear, and in the other half the immediate learning goals were clear.

The four treatments involved were:

Treatment 1. Direct teacher influence: clear goals, 35 dependent-prone students

Treatment 2. Direct teacher influence: unclear goals, 35 dependent-prone students

Treatment 3. Indirect teacher influence: clear goals, 35 dependent-prone students

Treatment 4. Indirect teacher influence: unclear goals, 35 dependent-prone students

The experimental population of 140 dependent-prone students was part of a larger group of 560 students. The larger group was exposed to the preceding four treatments in groups of 20 as part of a larger study. The 560 eighth grade students were selected at random from Minneapolis and St. Paul public schools. The 140 students were the top 25% in each treatment of the larger population according to their scores on the dependence-proneness test.

The students were brought, 20 per session, to a spare room in a public school. First, a pretest of geometry achievement and a test of dependence-proneness were administered. Second, a tape recording was

<sup>1</sup>This article is based on the PhD thesis of the first author. It was an adjunctive study of a larger project, the latter supported by a grant from the United States Office of Education.

played introducing the basic concepts of  $C = D$ , Distance = Speed  $\times$  Time, and geometric concepts and formulae concerning inscribed angles. In half of the groups the immediate goals were made clear because the recording explained how this information could be used to solve problems; in the other half, the goals were less clear because the students were warned that they could not be sure how this information could be used. Third, in the direct treatments a teacher gave a 15-minute lecture, with a few questions, explaining the material and illustrating problems that could be solved. In the indirect treatments, a teacher conducted a 15-minute discussion explaining the material and illustrating problems that could be solved. The content coverage was the same in the contrasting treatments. Fourth, the students then had about 15 minutes to practice solving problems at their seats by working on a problem sheet. And fifth, the posttest of achievement was administered. The entire sequence lasted 2 hours.

At appropriate points in this procedure the students' perceptions of goal clarity and their perceptions of the teacher were measured by paper-and-pencil tests. The reliability of both of these scales was estimated by the use of the Hoyt-Stunkard (1952) analysis of variance technique. The estimated reliability of the student perception scale was .64, while the reliability found for the goal perception measure was .93 for the measure administered just before the discussion (Amidon, 1959). Whenever the teacher talked, an observer classified all teacher and student statements according to Flanders' (1960) system of interaction analysis. Later the validity and reliability of the observer's judgments were verified by studying the tape recording that was made of every experimental session.

## RESULTS

### *Control of Teacher Influence and Goal Perception*

The manipulation of direct and indirect teacher influence occurred right after the tape recording when the teacher first came in contact with the students. The same role player acted as teacher in all treatments to avoid differences in personality and appearance differences between the distance. The differences between the direct and indirect approach are shown

TABLE 1  
PERCENTAGE OF TALLIES IN INTERACTION CATEGORIES

Category definition	Treatment			
	1	2	3	4
Teacher talk:				
Praise and encouragement	1.35	1.61	17.04	14.90
Clarification and development of ideas suggested by students	2.48	0.92	15.78	16.10
Asks questions	2.58	1.73	28.07	30.04
Gives own opinions and facts (lectures)	63.10	61.40	13.52	15.97
Gives directions	8.67	10.36	0.28	0.27
Criticizes students	13.03	15.54	1.27	0.94
Student talk	5.07	5.29	16.47	17.17
No one talking	3.49	3.45	7.75	4.69
Total tallies on which the percentage figures are based	889	869	711	746

in Table 1 according to the percentage of statements classified into interaction categories.

The figures in Table 1 show that essential differences between the direct and indirect treatments are: the teacher lectures and gives more directions in the direct treatments; he asks more questions and gets more student participation in the indirect treatments; he praises, encourages, and clarifies student ideas more frequently in the indirect treatments; and he criticizes students more frequently in the direct treatments.

The fact that the teacher controlled his behavior successfully and created the two teacher styles is self-evident. A Darwin (1959) chi square analysis of these same interaction data, after they were tabulated in a matrix of sequence pairs, was calculated to test the null hypothesis that there is no difference between interaction data of the direct and indirect combined treatments. The chi square value found was 702.2 ( $df = 90$ ). This value,

TABLE 2  
STUDENT PERCEPTION OF  
TEACHER INFLUENCE

Group	Mean	$F_{obs}$ between direct and indirect groups
Direct influence	11.74	
Indirect influence	18.84	78.40

transformed to a  $z$  score of 24.1 indicates that the differences could have occurred by chance with a frequency of much less than .01.

The reliability of the observers who classified the statements in the live situation was higher than .90. It can be shown that the errors of observation, at this level of reliability, are extremely small compared with the difference shown in Table 1 between the direct and indirect patterns.

After the lectures and discussions and before the period devoted to work sheets, the students responded to a number of opinion items combined into a scale which measured their perceptions of the teacher's behavior. An analysis of variance was made of these scale scores which indicated that the  $F$  ratio between the groups subjected to the direct and indirect influence treatments was 78.4 ( $df = 1/136$ ). This was significant at the .01 level of confidence. The mean scores indicated that students in the direct treatments more often saw the teacher as: "telling us what to do," "firm and businesslike," "making plans for us," "critical of our ideas," "talking more than the students," and "not using student ideas or suggestions." In the indirect treatments, students often marked: "finding out what we know," "relaxed and cheerful," "letting us make our own plans," "letting students talk," "using our ideas," and similar perceptions that were the opposite of the direct pattern. The

means and  $F$  ratio are presented in Table 2.

As expected, an analysis of observation data and of student perceptions of the teacher comparing the clear and unclear goal treatments showed no significant ( $p < .05$ ) difference in teacher behavior. Also, no significant ( $p < .05$ ) interaction effects were found in the analysis of variance.

The interaction analysis data and measures of the students' perceptions of the teacher did show that the differences between the direct and indirect approach did exist as required by the experimental design, were clearly seen by a trained observer, and noticed by the students.

Paper-and-pencil measures of the clarity of goals were made after the playback of the tape recording introducing the basic geometric concepts. Students responded to items such as: "Can you see all the steps necessary to finish your work?" "Right now can you see what you will be doing clearly?" "Can you picture your work so clearly that you could tell when you will be finished?" and similar items that were combined into a scale. An analysis of variance of these scale scores yielded an  $F$  ratio of 16.98 ( $df = 1/136$ ,  $p < .01$ ). An inspection of the means indicated that the results were consistent with the intended goal manipulation. The results show that the immediate response of students to the clear and unclear tape recording was significantly different. The means and  $F$  ratio are found in Table 3.

#### *Results of the Geometry Achievement*

Since achievement in geometry was the fundamental outcome variable analyzed in this study, the postachievement test was subjected to several analyses. The first analysis was the comparison of postachievement

scores between the indirect and direct teacher influence groups. The  $F$  ratio found in this analysis was 7.67, which was significant at the .01 level. The means of the indirect teacher influence groups were significantly higher than the means of the direct teacher influence groups on the postachievement measure. In order to reduce unaccounted for error in the analysis of postachievement, two analyses of covariance controlling preachievement scores and intelligence scores were run. The  $F$  ratio between direct and indirect teacher influence groups was 10.03 when intelligence was controlled, and 9.62 when pretest scores were controlled. Again, those  $F$  ratios were significant at the .01 level, indicating the superiority of the indirect teacher influence group. For each of the analyses of covariance  $df = 1/137$ .

The analysis of variance and covariance just discussed yielded insignificant ( $p > .05$ ) results when the clear and ambiguous goal perception groups were compared statistically. The interaction of goal perception and teacher influence also did not yield significant ( $p < .05$ ) results. The  $F$  ratios found in these analyses are presented in Table 4.

The means of the indirect group were significantly higher than the means of the direct group. This is true when intelligence and preachievement were controlled, and it was also true when they were not controlled. The means of the direct and indirect

TABLE 3  
STUDENT PERCEPTION OF THE GOAL

Group	Mean	$F_{obs}$ between clear and unclear groups
Clear goals	21.40	16.98
Unclear goals	26.13	

TABLE 4  
ANALYSIS OF DIFFERENCES IN THE MEANS  
OF POSTACHIEVEMENT SCORES WITH  
VARIOUS FACTORS CONTROLLED  
STATISTICALLY

Source of variation	$df$	$F_{obs}$	Null hypothesis
With no measure controlled:			
Interaction	1	—	Not rejected
Teacher influence	1	7.67	Rejected
Goal perception	1	—	Not rejected
Preachievement controlled:			
Interaction	1	2.71	Not rejected
Teacher influence	1	10.03	Rejected
Goal perception	1	—	Not rejected
Intelligence controlled:			
Interaction	1	1.08	Not rejected
Teacher influence	1	9.62	Rejected
Goal perception	1	1.13	Not rejected

TABLE 5  
MEANS OF DIRECT AND INDIRECT GROUPS  
FOR POSTACHIEVEMENT MEASURE

Mean	Not adjusted	Adjusted by pretest scores	Adjusted by intelli- gence
Direct group	9.24	9.31	9.30
Indirect group	10.82	10.76	10.69

teacher influence groups are presented in Table 5.

## DISCUSSION

The measures of geometry achievement indicate that the dependent-prone students learned more in the classroom in which the teacher gave fewer directions, less criticism, less lecturing, more praise, and asked more questions which increased their verbal participation. This finding takes on added significance when compared

with Flanders' (1960) findings that the total group of 560 students, those scoring high, average, and low on the dependence-proneness tests, failed to show the same significant differences under the same conditions. Moreover, when the independent-prone students (those in the lower 25% on the dependency scale) were compared separately, no differences were found. Compared with students in general, dependent-prone students are apparently more sensitive to the influence pattern of a geometry teacher.

The authors are disposed to interpret these findings in terms of the probable effects of teacher influence on the dependent-prone student. We assume that dependent-prone students are more sensitive to the directive aspects of the teacher's behavior. As the teacher becomes more directive, this type of student finds increased satisfaction in more compliance, often with less understanding of the problem solving steps carried out. Only when he is free to express his doubts, to ask questions and gain reassurance, does his understanding keep pace with his compliance to the authority figure. Lacking this opportunity, compliance alone may become a satisfactory goal and content understanding may be subordinated to the process of adjusting to teacher directives. It is interesting to note that this effect occurs even when the material being learned concerns an orderly, logical system, such as exists in geometry.

One implication of this study is that closer supervision through the use of direct influence, an all too common antidote to lower achievement, may be more harmful than helpful for dependent-prone students.

#### SUMMARY

The primary purpose of the study described here was to determine the

effects of direct vs. indirect teacher behavior and of clear vs. unclear student perception of the learning goal on the achievement of eighth grade geometry students. A specially trained teacher role played both a very direct and a very indirect teacher in a laboratory situation involving 140 eighth grade pupils chosen from a larger population on the basis of high scores on a test of dependency proneness. All students were randomly assigned to one of the following four experimental treatments: direct teacher influence with clear goals, direct teacher influence with unclear goals, indirect teacher influence with clear goals, and indirect teacher influence with unclear goals.

Students in the various classifications were then compared on the basis of pre- and postachievement tests in geometry. No differences were found between the clear goal and unclear goal treatments, indicating that in this study, at least, achievement of dependent-prone students was not affected by perception of the learning goal. An analysis of the direct and indirect treatments indicated that the children taught by the indirect teacher learned more than did the children taught by the direct teacher.

The results of this study take on additional meaning when compared with the results of Flanders (1960) using the same experimental design. Flanders found no differences (among the four experimental conditions) on the total group of 560 students who ranged from very high to very low on the dependence scale. Apparently dependent-prone students are more sensitive to types of teacher influence than are independent-prone students or students who make average scores on the test for dependence proneness.

## REFERENCES

- AMIDON, E. J. Dependent-prone students in experimental learning situations. Unpublished doctoral dissertation, University of Minnesota, 1959.
- ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh; Carnegie Press, 1951.
- DARWIN, J. H. Note on the comparison of several realizations of a Markoff chain. *Biometrika*, 1959, **46**, 412-419.
- FLANDERS, N. A. Teacher influence on pupil attitudes and achievement. Final Report; 1960; University of Minnesota; Project 397; United States Department of Health, Education, and Welfare; Cooperative Research Program; Office of Education.
- FLANDERS, N. A., ANDERSON, J. P., & AMIDON, E. J. Measuring dependence proneness in the classroom. *U. Minn., Bur. Educ. Res. res. Memo.*, 1960, No. Ber 60-6.
- HOYT, C., & STUNKARD, C. L. Estimation of test reliability for unrestricted item scoring methods. *Educ. psychol. Measmt.*, 1952, **12**, 756-758.
- KAGEN, J., & MUSSEN, P. Dependency themes on the TAT and group conformity. *J. consult. Psychol.*, 1956, **20**, 19-27.
- LIVSON, N., & MUSSEN, P. The relation of control to overt aggression and dependency. *J. abnorm. soc. Psychol.*, 1957, **55**, 66-71.
- SMITH, D. E. P. Fit teaching methods to personality structure. *High sch. J.*, 1955, **39**, 167-171.
- WISPE, L. C. Evaluating section teaching methods in the introductory course. *J. educ. Res.*, 1951, **45**, 161-186.

(Received March 13, 1961)

# THE INFLUENCE OF INTERRACIAL CONTACT ON SOCIAL ACCEPTANCE IN A NEWLY INTEGRATED SCHOOL<sup>1</sup>

STATEN W. WEBSTER<sup>2</sup>

*University of California, Berkeley*

Does interracial contact lead to improved attitudes and increased social acceptance among the participants? This question has been the subject of a number of investigations, most of which have involved white and Negro adult subjects.

Data reported in these studies have been contradictory. A number of investigations (Brophy, 1945; Deutsch & Collins, 1952; Neprash, 1952; Wilner, Walkley, & Cook, 1952) have indicated the successful influence of interracial contact in reducing prejudice. However, a number of negative findings have also been reported (Mussen, 1950; Rose, 1947; Young, 1932).

Those who argue against the positive effects of contact cite the restrictive influence of racial and ethnic group norms upon individuals in such situations. The existence of such frames of reference has been clearly documented by Bogardus (1959). While the impact of racial or ethnic group membership upon the individual has been established, there is also evidence that under certain conditions individuals and groups do modify their interracial attitudes, feelings, and behaviors. An analysis of successful contact situations seems to reveal three important requisite factors: equal status of the participants in the situa-

<sup>1</sup>This study is abstracted from S. W. Webster's doctoral dissertation written in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of California, Berkeley, 1960.

<sup>2</sup>Now Associate Research Director of the project, Studies in the Dynamics of Administrative Interaction, at the School of Education, University of California, Berkeley.

tion, contact of sufficient duration and intimacy, and the sanction or support of a higher authority.

## PROCEDURE

### Problem

In view of the dearth of contact research involving public school settings, this investigation sought to study the effects of interracial contact in a newly integrated northern California junior high school.

Specifically, two major questions were investigated: Does interracial contact, over a 6-month period, between white and Negro subjects, lead to increased expressions of social acceptance of one group for the other? Are precontact social acceptance scores (as measured by a social acceptance scale) associated with the eventual making of interracial friendships?

*Setting.* The school board of a high school district in the San Francisco Bay Area reached a decision in the spring of 1959 to effect certain changes in the boundaries surrounding its junior and senior high schools. These realignments were made necessary by overcrowded conditions in some schools, and the unequal distribution of children in terms of socioeconomic status.

Prior to the boundary modifications, which became effective in September 1959, Jackson Junior High School<sup>3</sup> was populated by an all white student body. The composition of the school population was altered greatly by the boundary changes, with some 300 Negro children being integrated into the school's present total of some 1,400 students. This school, its feeder elementary schools, and its surrounding communities served as the setting for this study. While there were some socioeconomic differences between the white and Negro communities, a study of the personal histories of the children revealed these to be minor in nature.

*Subjects.* When the new school boundaries were announced late in the spring, it became

<sup>3</sup>Actual school names are not used in this report.

obvious that most children in two elementary schools would be attending, in the fall, a truly integrated school for the first time. Walker Elementary School, which was all white, sent 60 sixth grade graduates to Jackson Junior High in September. From the almost all Negro Newell School came 44 transfers. (This school contained, in addition to some 400 Negro students, 17 children of Indian and Mexican-American extraction.)

A socioeconomically similar control group comprised of 55 whites and 53 Negroes was also selected. An all white community and junior high school, in a neighboring city, provided the white control subjects. The Negro control subjects were selected from an almost all Negro junior high school in another nearby city; with the exception of 9 Mexican-American and white children, the latter school's student population of 450 was composed entirely of Negroes.

*Hypotheses.* Four hypotheses were investigated in this study; they were as follows:

Hypothesis I: White and Negro subjects who have experienced intergroup contact in a school situation will express significantly more social acceptance of each other after a 6-month period of time.

Hypothesis II: The postcontact social acceptance scores of the white and Negro experimental subjects will display significantly more out-group acceptance than will the non-contact scores of the control subjects.

Hypothesis III: The precontact social acceptance scores of the experimental subjects will be significantly associated with their later friendship choices.

Hypothesis IV: The retest social acceptance scores of the experimental subjects will be significantly associated with their interracial friendship choices.

These hypotheses are interrelated. Hypotheses I and II seek to measure the impact of contact upon the experimental subjects, in light of the data on the control subjects. Hypotheses III and IV are designed to seek additional data on the effects of contact and on the validity of the social acceptance scale used to gather these data.

### Method

The test-retest design was deemed most appropriate for this type of study. Both experimental groups were tested initially in June 1959, while they were still attending their particular elementary schools, and were retested at Jackson School in March 1960. Because of the time factor, it was impossible

to test the control groups in spring 1959. The failure to obtain these data is, admittedly, a weakness of the study. However, since their elementary and junior high schools were relatively homogeneous in racial composition, it seems valid to assume that their single test scores (spring 1960) on a social acceptance scale (see below) would not differ significantly from the precontact scores of their matched experimental groups.

All testing sessions were conducted by members of the subject's racial group. To avoid the possibility of inhibiting the responses of the children, they were told that the questionnaires were anonymous. However, although the experimenter entertained some reservations about the procedure, invisible ink markings were used on the social acceptance scale to identify the respondents.

To measure social acceptance, a nine-item social acceptance scale was constructed. The respondent was asked to rank his willingness to accept members of four groups (Chinese, Japanese, Mexicans, and whites or Negroes) in nine social situations. The social situations ranged from the less intimate (e.g., working in the same factory) to the very intimate (e.g., living next door). Scale scores from 0 to 9 were possible, with the higher scores representing the greater acceptance. The scalability of the instrument was established in keeping with the criteria suggested by Guttman (1950). The coefficient of reproducibility was .92.

In addition to the social acceptance scale, a brief sociometric friendship questionnaire was used. The subjects were asked to indicate their six best friends in the school and to give the ethnic or racial backgrounds of these friends. These data were obtained after the retest of the experimental subjects.

To test the above hypotheses, it was necessary to validate the assumption that both experimental and control groups were initially similar in levels of social acceptance. The two-tailed application of the Kolmogorov-Smirnov two-sample test (Siegel, 1956) was used in this analysis. Since the KS test is sensitive to differences in central tendency, dispersion, and skewness, it seemed especially appropriate for this purpose. The comparison of the experimental subjects' precontact scores and the single test scores of the control subjects revealed no significant differences.

### RESULTS

Table 1 contains a summary of the basic test data on the experimental

TABLE 1  
TEST AND RETEST SCORES OF EXPERIMENTAL  
AND CONTROL GROUPS ON A SOCIAL  
ACCEPTANCE SCALE

Subjects	Group	N	Range	$\bar{X}$	Mode	SD
<b>Experimental:</b>						
Pretest (June 1959)	White	60	0-9	1.85	0	2.32
	Negro	44	0-9	6.05	9	2.86
<b>Posttest</b> (March 1960)						
	White	60	7-0	1.48	0	2.02
	Negro	44	9-0	6.32	9	2.61
<b>Control:</b> (March 1960)						
	White	55	0-9	2.54	0	2.24
	Negro	53	0-9	5.98	6	3.97

and control groups. The reader will note the great discrepancies between the scores of the Negroes and the whites. Bogardus (1959) has reported similar distributions in one of his studies of social distance.

*Hypothesis I.* The Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) was used to test the first hypothesis.<sup>4</sup> This nonparametric test makes use of comparisons between pairs of scores and takes into consideration both the magnitude as well as the direction of the differences between the sets of scores. Change for the white subjects was in the negative direction and significant ( $z = 3.96, p < .05$ ). Change for the Negro subjects was positive and significant ( $z = 2.12, p < .05$ ). Thus, the predicted hypothesis was confirmed only in the case of the Negro subjects.

An equal number of Negro subjects changed in each direction. However, the magnitude of change was greater for those subjects whose gains were positive.

*Hypothesis II.* The testing of Hypothesis II involved comparing the postcontact scores of the experimental

groups with those of their controls, in an attempt to see if the contact situation had produced significant differences among the groups.

The one-tailed application of the KS test revealed a significant difference between the scores of the white experimental and control groups ( $\chi^2 = 7.23, df = 2, p < .05$ ); however, this difference was in the nonpredicted direction. The obtained difference between the scores of the Negro experimental and control subjects was not significant ( $\chi^2 = 2.16, df = 2, p > .05$ ).

*Hypothesis III.* This hypothesis is based on a suggestion by Neprash (1952). He speculated that the making of out-group friendships in a contact situation may be simply a matter of favorable precontact attitudes toward that group and not a result of contact. To test this assumption the precontact scores of the experimental subjects were dichotomized at the median and a comparison was made of the number of out-group friendships indicated by each group. The  $\chi^2$  test was used in the analysis and the findings were not significant. The median, in this case, happened to fall close to the derived zero intensity point (discussed later).

These findings seem to indicate that, in this study, precontact attitudes played little or no part in the willingness to make out-group friendships in the contact situation. The making of interracial friendships, it appears, was related to the effects of contact itself.

*Hypothesis IV.* It was predicted in Hypothesis IV that the subjects' retest social acceptance scores would be significantly associated with their reported interracial friendships. In testing this assumption, use was made of the intensity measure suggested by Guttman (1954) and his associates.

While the method of deriving this score will not be fully discussed, a

<sup>4</sup>The .05 level of significance was used in testing all hypotheses.

brief description will be given. The zero or low intensity point is a measure that seeks to establish a psychologically meaningful point of low intensity on an attitude scale. Through the use of the extreme responses of the subjects to each scale item (Answer Categories 1 or 4), it is possible to trichotomize the respondents into three general groups: intensely positive subjects, intensely negative subjects, and indifferent or low intensity subjects.

In the case of the social acceptance scale used in this study, the zero intensity point was found to be located at the scale score of four.

The subjects were dichotomized on the basis of their intensity scores, and an analysis was then made of the number of out-group friendship choices reported by each group. The obtained result was significant ( $\chi^2 = 10.56$ ,  $df = 1$ ,  $p < .01$ ) and Hypothesis IV, was thus confirmed.

### DISCUSSION

The positive effects presumably associated with interracial contact were not confirmed in this study. What, then, are some factors which seem to account for the obtained results?

Although the three-part criterion presented earlier in this paper was met in this contact situation, it is possible that only more general and intimate contact between interracial groups (Brophy, 1945; Deutsch & Collins, 1952; Wilner et al., 1952) results in significant modifications of attitudes. It should be remembered that the subjects, while attending the integrated school, continued to live in racially closed communities. It is possible that integration within the total social environment is necessary, for the attitudes and values which can be developed in the school setting are probably contradicted by the realities of the restricted community.

The length of the contact situation in this instance could be another factor that partially explains the failure of the contact hypotheses. A 6-month period of time could very well be inadequate to allow for the necessary intergroup adjustment of the subjects. Although this period of contact was longer than that reported in a number of studies, subsequent investigation may show that the initial assumptions associated with contact experiences are valid.

The social acceptance scores of both white and Negro subjects provide another possible explanation. Social distance (acceptance) research by Bogardus (1959) and others has documented the existence of a hierarchy of racial preferences. These normative frames of reference appear to be held both commonly and consistently by members of various racial or ethnic groups. Traditionally, whites rate Negroes in their preferences low on a social distance scale, while Negroes tend to rate whites quite highly. The data obtained in this study reflect these phenomena. While social distance measures have proven to be both reliable and valid, it is possible that they are not sensitive enough to measure less noticeable changes within individual attitude structures. Methods and instruments which assess interracial attitudes with greater depth and precision would seem to be needed.

An interesting finding of the study was the significant association of interracial friendship choices with post-contact social acceptance scores but not with pretest scores. This finding seems to suggest caution in accepting the responses of subjects to social distance and other types of attitude measures as predictors of behavior. Many subjects whose precontact responses were favorable or highly accepting of the other group failed to

make eventual interracial friendships, and in a number of cases completely reversed their scale scores from high acceptance (scores of 7-9) to high rejection (0-2).

The data obtained in this study seem to suggest a critical reappraisal of the contact hypothesis. Factors external to the setting of the interaction as well as those associated with racial and ethnic group membership must be considered.

### SUMMARY

This study tested three major hypotheses relating to interracial contact, in a California junior high school, between white and Negro subjects. It was held that (a) a school contact situation of 6-months' duration would lead to increased interracial social acceptance by members of the two racial groups, (b) that a comparison between experimental subjects who experienced interracial contact and similar control subjects who did not would reveal greater interracial acceptance among the former group, and (c) that precontact social acceptance scores of the experimental subjects would be associated with out-group friendship choices made in the contact setting.

The test-retest design was used with 60 white and 44 Negro subjects, and 55 white and 53 Negro control subjects. A single administration of the social acceptance scale took place with the control subjects. All tests were administered by adults of the particular subjects' racial groups.

Contact had a negative effect upon the white subjects; they became significantly less accepting of Negroes. The findings were inconclusive in the case of the Negro subjects, but did

tend to indicate that change was greater in the direction of more acceptance of whites. Precontact social acceptance scores were not found to be associated with later friendship choices. However, postcontact acceptance scores were associated with actual out-group friendship choices at the .01 level of confidence.

### REFERENCES

- BOGARDUS, E. Race reactions by sexes. *J. Sociol. soc. Res.*, 1959, 43, 439-441.
- BROPHY, I. The luxury of anti-Negro prejudice, *Publ. opin. Quart.*, 1945, 9, 456-466.
- DEUTSCH, M., & COLLINS, MARY E. The effects of public housing upon interracial attitudes. In G. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (Rev. ed.) New York: Holt, 1952. Pp. 582-593.
- GUTTMAN, L. The basis of scalogram analysis. In S. Stauffer, E. A. Suchman, P. F. Lazarsfeld, Shirley A. Star, & J. A. Clausen (Eds.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 60-90.
- GUTTMAN, L. Principal components of scalable attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. Glencoe, Ill.: Free Press, 1954. Pp. 229-233.
- MUSSEN, P. H. Some personality and social factors related to changes in children's attitudes toward Negroes. *J. abnorm. soc. Psychol.*, 1950, 45, 423-441.
- NEPRASH, J. Racial group contacts and social distance. *Phylon*, 1952, 16, 207-212.
- ROSE, A. M. *Studies in reduction of prejudice*. Chicago: American Council on Race Relations, 1947.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- WILNER, D. M., WALKLEY, R. R., & COOK, S. W. Residential proximity and intergroup relations in public housing projects. *J. soc. Issues*, 1952, 8, 45-69.
- YOUNG, D. *American minority peoples*. New York: Harper, 1932.

(Received April 3, 1961)

## TEACHERS' PERSONALITY RATINGS OF HIGH SCHOOL PUPILS

H. J. HALLWORTH

*Department of Education, University of Birmingham, England*

Recent contributors to this journal have indicated the need for research into teachers' ratings of their pupils, a need which is equally pressing in Europe as in America. The present article describes part of such a research which was carried out from the Department of Education in the University of Birmingham, England. It therefore concerns a sample drawn from a different educational tradition from those studied, for example, by Ryan (1958) and Holland (1959). The results, however, are remarkably comparable to theirs and, with one exception, discrepancies appear to be due entirely to differences in measuring instruments and statistical procedures.

It is hoped that comparative data such as this will be of interest to American workers both in affording confirmation of their findings and in suggesting small differences which may prove illuminating.

### METHOD

#### Subjects

At the age of 11+ years most English children pass from primary to secondary schools. Approximately 20% with the highest intelligence and attainment are allocated to grammar schools; most of the remainder pass into secondary modern schools.

Ratings were obtained for boys in first year (11+ years) and fourth year (14+ years) classes of six single-sex secondary schools in an industrial region near Birmingham. The schools were selected in three matched pairs, each pair comprising a boy's grammar school and a boy's secondary modern school within one locality. The subjects were therefore divided according to age and school type into four groups, each distributed over three schools and containing approximately 112 boys.

It is commonly accepted that there are sex differences in certain personality traits: boys

are assumed to be more self-assertive, and girls more cooperative with adults. Such differences have frequently been demonstrated in mean scores obtained on personality inventories (e.g., Castaneda, McCandless, & Palermo, 1956). Groups comprising members of both sexes may consequently be expected to yield spuriously high correlations between certain personality traits. It was therefore considered necessary to obtain ratings for groups consisting of pupils of like sex only. However, in the complete research of which part is reported in this article, ratings were obtained for parallel groups of girls and yielded results comparable to those given below.

Similarly, it may be claimed that some personality traits are more typical of one age than of another. Ratings obtained for a group with a wide age range may therefore have spuriously high intercorrelations. It was therefore considered necessary to obtain ratings for several classes of boys all of the same age level, and to combine these into one group. This was done for two age levels and for two types of school in order to determine the effect of age and school type upon the intercorrelations obtained.

#### Measures

Each class teacher rated his pupils on 12 personality traits, using a graphic five-point scale. In order to avoid spuriously high intercorrelations from another source, he was requested to complete all ratings on one trait at a time. For each trait there was provided a brief description for each of the extreme ends of the scale, and a two-word description of each of the five points on the scale. The traits were as follows:

1. Emotional stability
2. Trustworthiness
3. Persistence
4. Cooperation with teachers
5. Cheerfulness
6. Sense of humor
7. Sociability
8. Self-assertion-submission
9. Maturity
10. Popularity
11. Confidence-anxiety
12. Spontaneity-withdrawal

When several teachers rate different children and the ratings on different traits are then intercorrelated, there is yet another possible source of spurious correlation: some teachers give high ratings on all traits, some give low ratings. In the present research therefore, no attempt was made to use ratings made by different teachers when obtaining correlations between traits. It was accepted that neither mean nor standard deviation could be assumed to be comparable for ratings made by different teachers or, indeed, for ratings made by one teacher on different traits. Teachers were therefore requested, when making ratings on any one trait, to take the mean of their own class as the mean of their rating scale. They were also requested to distribute the ratings according to the percentages of the normal curve, a graphic model being provided for this purpose.

### Analysis

The data was analysed with the aid of an electronic computer. A correlation matrix was first obtained for the ratings made on the 12 traits by each teacher for the boys in his own class. Within each of the four groups the correlation matrices were then combined, using  $z$  scores, to obtain a matrix of average intercorrelations between traits for each group. A principal components analysis was then performed on each of the four combined matrices.<sup>1</sup>

### RESULTS

Across each of the four analyses the first two components were essentially similar and accounted for approximately 60% of the total variance. Further components invariably accounted for less than 10% of the total variance and, since they had little similarity across the analyses, will not be considered in this article. There is general agreement regarding the large error variance in teachers' ratings. However, when comparable factors are obtained from four different analyses,

this suggests a certain regularity in the manner in which ratings are made and justifies further examination.

The first two components for each of the four analyses are shown in the accompanying diagrams, Component I being shown along the  $x$  axis and Component II along the  $y$  axis. The percentage of the total variance extracted by each component is indicated.

In the diagram for first year grammar school boys, a group of ratings falls close to each of the main axes. Those falling near the axis for Component I are cheerfulness, sense of humor, sociability, self-assertion, and spontaneity, each with approximately 64% of its variance accounted for by this one dimension. The ratings falling approximately along the axis for Component II are emotional stability, trustworthiness, persistence, cooperation with teachers, and maturity, each with between 38% and 66% of its variance accounted for by this dimension. We may name the first dimension "extraversion" and the second "reliability and conscientiousness." Two other ratings, namely popularity and confidence, lie close together and between these dimensions.

The diagram for fourth year grammar school boys has the same two clusters of ratings, albeit they are in this case less compact. A slight orthogonal rotation would bring them onto the main axes. Again the ratings on confidence and popularity lie together and between the two clusters.

The two diagrams for secondary modern school boys appear at first sight to show the well-known "halo-effect" which occurs when ratings on a number of traits are made of an individual. All ratings have a heavy loading on Component I. For first year boys, however, the same two clusters may be seen, although the extraversion traits are more widely spread. In the analysis for fourth year boys, one

<sup>1</sup> Acknowledgements are due to W. L. B. Nixon, of the University of London Computer Unit, whose Mercury computer principal components (analysis) program was used. The program for the combination of correlation matrices, using  $z$  scores, was developed by the author in conjunction with the Computer Unit.

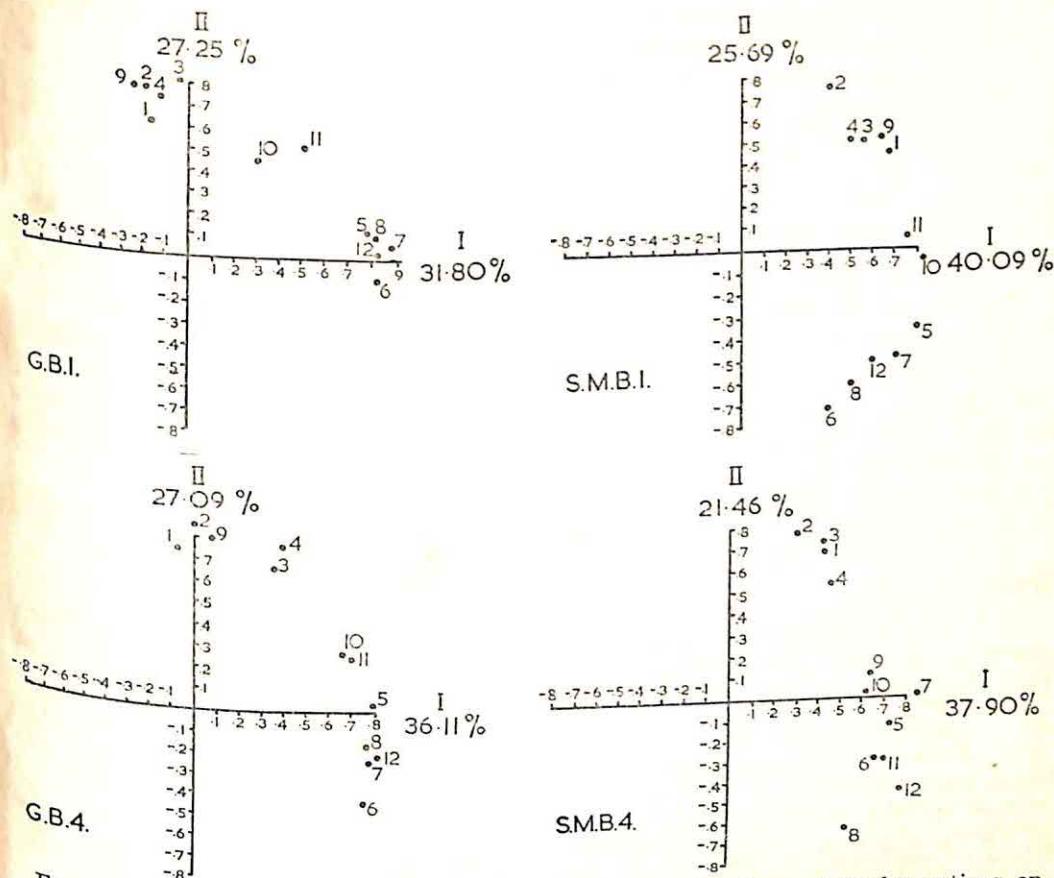


FIG. 1. Principal components analyses for four groups: Loadings of twelve ratings on two components. (I = first component, II = second component. The percentage of variance extracted by each component is shown. GB<sub>1</sub> = grammar school boys Year 1, GB<sub>4</sub> = grammar school boys Year 4, SMB<sub>1</sub> = secondary modern school boys Year 1, SMB<sub>4</sub> = secondary modern school boys Year 4. 1 = emotional stability, 2 = trustworthiness, 3 = persistence, 4 = cooperation with teachers, 5 = cheerfulness, 6 = sense of humor, 7 = sociability, 8 = self-assertion-submission, 9 = maturity, 10 = popularity, 11 = confidence-anxiety, 12 = spontaneity-withdrawal.)

cluster is made by the traits of trustworthiness, persistence, emotional stability, and cooperation. The extraversion traits are again scattered and now include among them the ratings of confidence, popularity, and maturity.

#### DISCUSSION

##### Grammar and Secondary Modern School Groups

It is possible that the greater degree of apparent halo effect for the two secondary modern school groups is due to psychologically irrelevant factors in the data or in the method of analysis.

Further research is necessary to determine the matter, and is currently in progress. A consistent difference related to school type appears, however, to demand a psychological explanation.

There is no reason to suppose that the secondary modern school teachers had less acquaintance with their pupils and therefore less opportunity to differentiate their ratings. Nor is there any reason to assume that their ratings were in any other way made under less favorable conditions.

It would appear, therefore, that if

any psychological explanation is required of the difference between the two sets of analyses, it is probably to be found in the way in which the two groups of teachers "perceived" or thought about their pupils.

### *Interpretation of the Principal Components*

Ryan (1958) reports high intercorrelations between all the traits which he used and concludes that "teachers rate on the basis of general impression." This is, admittedly, a common and reasonable conclusion. It is not, however, an invariable finding, as is shown by the above diagrams for grammar school boys, where little halo effect is found.

The interpretation of the two principal components for each group of grammar school boys is quite clear. In neither case is there any appreciable halo effect. Two dimensions, one of extraversion, one of reliability and conscientiousness, account for three-fifths of all the variance.

The problem arises when we consider the secondary modern school groups.

If we interpret the two principal components as they stand, we shall label Component I "halo effect." Component II is, however, difficult to name, and here lies the difficulty: It must be described in terms of a scale running from cooperation, trustworthiness, persistence, and emotional stability at one end, to sociability, cheerfulness, spontaneity, self-assertion, and sense of humor at the other. Such a scale does not make good psychological sense. Moreover, the suggestion that the clusters which we have named "reliability and conscientiousness" and "extraversion" are independent dimensions for grammar school groups, but opposite ends of a single scale for secondary modern school

groups, does not make good common sense.

At this point we may note that the factorial solution provided by the components analysis is only one of many possible solutions. It has already been indicated that a small rotation is appropriate for the analysis of the fourth year grammar school group. It is now suggested that the obvious solution for the secondary modern groups is to rotate the two components orthogonally until the apparent halo effect is at a minimum and we have the two dimensions already obtained for the grammar school groups.

That this is practicable may readily be seen from the two diagrams. The extraversion dimension is indeed not as clearly defined as could be desired, and it is possible that a still better description would be in terms of a circumplex model (Schaefer, 1959). The rotation does, however, produce better psychological meaning in the analyses for the secondary modern groups, and makes the analyses for all four groups comparable.

### *Traits and Dimensions*

If we accept this solution, there remains to consider the difference between the present research and certain American work, such as that already quoted, which has been interpreted in terms of halo effect only. The difference is almost certainly due in part to different procedures employed in obtaining ratings. Table 1 gives the correlation matrices for the two first year groups, and it is readily apparent that the correlations are lower than those obtained by either Ryan or Holland. The most significant difference lies, however, in the nature of the traits used. Ryan, for example, used eight traits: seriousness of purpose, industry, initiative, influence, concern for others, responsibility, self-control

TABLE 1  
THE CORRELATION MATRICES FOR TWELVE TEACHER RATINGS FOR GRAMMAR SCHOOL BOYS  
YEAR 1 AND SECONDARY MODERN SCHOOL BOYS YEAR 1

Trait	1	2	3	4	5	6	7	8	9	10	11	12
1. Emotional stability		46	40	27	-03	-15	-08	-08	40	09	37	-23
2. Trustworthiness	55		56	53	-03	-16	-16	-13	49	28	21	-15
3. Persistence	54	55		57	12	-11	-06	09	58	21	32	02
4. Cooperation with teachers	51	59	37		04	-16	-01	-12	55	24	14	-02
5. Cheerfulness	43	11	25	29		71	58	45	-18	21	37	56
6. Sense of humor	-07	-35	-11	-13	67		57	58	-31	21	30	55
7. Sociability	19	00	10	23	75	56		66	-09	24	44	74
8. Self-assertion-submission	08	-28	02	-08	56	42	66		00	30	46	59
9. Maturity	61	62	58	49	30	-15	15	11		26	27	-06
10. Popularity	50	31	37	44	66	36	59	32	44		27	15
11. Confidence-anxiety	51	25	52	24	48	15	42	38	56	53		31
12. Spontaneity-withdrawal	21	-08	07	-05	56	49	57	59	18	42	46	

Note: Correlations for grammar school boys Year 1 are above the diagonal. Correlations for secondary modern school boys Year 1 are below the diagonal.

neatness and personal appearance. All would appear to fall on, or close to, the dimension of reliability and conscientiousness. The same comment applies, in general, to the 12 ratings used by Holland except, possibly, popularity, social leadership, and physical vigor. In neither case was any typical set of extraversion traits used. It would therefore be impossible to obtain two dimensions similar to those obtained in the present study, and the most comparable result would be the halo effect that was actually obtained in each case.

In the present research, ratings on the traits of popularity and confidence tend to fall together and between the two main dimensions. Here, possibly, is the beginning of a third dimension which is not, however, orthogonal to the other two. There is no reason to suppose that yet other dimensions could not be identified. It is, however, apparent that they would be of less importance both practically and theoretically. It is equally apparent that the dimension of reliability and conscientiousness, already so well identified in earlier articles in this journal, is

of greater importance to teachers than is the dimension of extraversion.

#### Ratings of Maturity

The rating of "maturity" is of particular interest in the present research, since in three analyses it is part of the dimension of reliability and conscientiousness, and in the analysis for fourth year secondary modern school boys it is closer to the extraversion traits. An interpretation of this difference appears to demand some acquaintance with secondary education in England.

Grammar schools in England are essentially academic schools and grammar school teachers value most highly in their pupils those traits which are associated with academic success. Both these teachers and, apparently, teachers of first year secondary modern school boys, regard as mature those pupils who are most trustworthy, persistent, emotionally stable and cooperative: in other words, the children who are most mature in their school work.

Teachers of fourth year secondary modern school boys associate maturity in their pupils more closely with those

traits which may be immediately useful in life outside school. Very few of these boys could be expected to remain at school beyond the end of their fourth year. It appears that their teachers were already assessing their maturity in terms of the social qualities they would require upon leaving school.

The implication is that ratings of maturity have a large "evaluative" element. It is probable that most teachers, as Holland found, measure maturity in relation to the dimension of reliability and conscientiousness. There appear to be, however, circumstances in which teachers relate maturity to other traits which they consider more relevant for their pupils.

### CONCLUSION

The results of this research indicate that the extent to which trait names are interchangeable depends upon the traits with which we are dealing. If we refer to those traits with which teachers are commonly concerned and which in this article have been designated as reliability and conscientiousness, then indeed they are interchangeable as Ryan suggests; to measure one of them, such as persistence or emotional stability, would be effectively to measure all. If, however, we also refer to such traits as sociability, then there is a case for maintaining that here is a quite different dimension. If still other traits were used, other dimensions again might appear. "Popularity" and "confidence" tend to be synonymous, and relatively independent of either of the two dimensions obtained in this research. "Maturity" changes its meaning, as if teachers always ask themselves: "Maturity for what?"

In so far as we obtain a halo effect we may say that when a teacher is rating one of his pupils on a number of traits he tends to ask himself one question only: "How does he get on with

me?" or "How much do I approve of him?" In effect, it is suggested in this article that, if we look more closely at ratings on a variety of traits, we may reasonably conclude that teachers ask themselves not one question, but two, namely, "How does he get on with me?" (reliability and conscientiousness); and "How does he get on with other boys?" (extraversion).

It is apparent that the answers to these two questions alone will go far to explain teachers' ratings on all the traits used in this research.

The question whether they may be described as referring to two teachers' stereotypes will be dealt with in a later article.

### SUMMARY

Teachers in six boys' secondary schools in England rated their pupils on 12 personality traits, and a correlation matrix was obtained for each teacher. Matrices were combined for classes of different age and school type into four groups. In contrast to the usual finding of a "halo" effect, two dimensions were obtained in each case and were identified as "reliability and conscientiousness" and "extraversion." It is suggested that when rating pupils on a variety of traits, teachers may be described as asking themselves two questions: "How does he get on with me?" and "How does he get on with other boys?"

### REFERENCES

- CASTANEDA, A., McCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Developm.*, 1956, 27, 317-326.
- HOLLAND, J. L. Some limitations of teacher ratings as predictors of creativity. *J. educ. Psychol.*, 1959, 50, 219-223.
- RYAN, F. J. Trait ratings of high school students by teachers. *J. educ. Psychol.*, 1958, 49, 124-128.
- SCHAEFER, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, 59, 226-235.

(Received April 7, 1961)

## CONCEPT ATTAINMENT AS A FUNCTION OF NUMBER OF POSITIVE INSTANCES PRESENTED

EUGENE SHORE  
*Pennsylvania State University*

AND

LEE SECHREST  
*Northwestern University*

Callantine and Warren (1955) have reported an experiment in which it was demonstrated that superior performance on a concept attainment problem results from the repeated presentation of one or a few instances of the concept as opposed to single presentations of a larger number of instances. However, superior performance in this case refers to errors made during learning, i.e., associating the concept name (nonsense syllable) to the instances presented. Callantine and Warren point out that performance on *new* instances, what they refer to as transfer, was better in the group which received single presentations of 20 different instances.

The additional research evidence bearing on the problem of concept attainment and variety of instances is not totally consistent. Luborsky (1945) using an aircraft recognition task found eight exposures of a plane to be more effective than three exposures, but the five additional exposures differed qualitatively from the common three, which were diagrammatic. Reed (1946) reported more "consistent" concepts when more instances were given, but he did not control for the number of times each instance was presented to each subject. And, Adams (1954), using a problem based on spatial arrangements of lights found, somewhat surprisingly, that training on a single kind of task was better than multiple task training in a transfer test series. Morrisett and Hovland (1959) repeated Adams' study with an additional group and concluded that both a variety of problems and a high

level of learning are necessary for transfer, and, presumably, concept learning to take place.

The present experiments were designed to extend the work of Callantine and Warren, who used only a single type of concept. Specifically, in the studies to be reported concepts based on nonsense syllables were chosen to represent differing levels of difficulty and differing concept attributes. The first experiment proved to involve possible methodological complications which were avoided in the second.

### EXPERIMENT I

#### *Procedure*

Sixty-four male subjects from an introductory psychology class were assigned randomly to one of four experimental conditions:

Group 3 was presented with three different instances of each of four concepts, each instance being presented six times.

Group 6 was presented with six different instances of each of four concepts, each instance being presented three times.

Group 9 was presented with nine different instances of each of four concepts, each instance being presented two times.

Group 18 was presented with 18 different instances of each of four concepts, each instance being presented only once. Thus, each of the four groups was presented with a total of 72 separate stimuli, 18 for each concept. The groups differed in terms of the number of different instances of the concepts which were to be discovered.

The concepts selected for study were derived from common characteristics of certain nonsense syllables. On the basis of preliminary study, concepts at four levels of difficulty were chosen for use in the study, difficulty having been defined in terms of the inverse of the number of test instances of the concept which could be identified

after a training series on each of the concepts.

The nonsense syllables were obtained from Glaze's lists (Hilgard, 1951), with an attempt being made to utilize only syllables of low association value. Most of the instances were chosen from the 13%, 20%, and 27% association value lists. Each of the four concepts to be utilized was assigned a simple name for use in the experiment. The definitions, names, and sample instances used for each of the concepts are presented in Table 1.

The subjects were run in small groups of from 6 to 10 members each. The instances of the concepts were printed on white 4 X 6 index cards. They were exposed by hand against a neutral background at the rate of one every 13 seconds. The subjects were given 8 seconds in which to mark an appropriate answer in a booklet, were told to

TABLE 1  
DEFINITIONS, NAMES, AND SAMPLE  
INSTANCES OF CONCEPTS

Concept name	Definition	Instances
DASH TRIANGLE	Middle letter A Ending in X, Y, or Z	KAG, ZAB, XAP, PAQ GOX, PIY, JEZ, KUY
DOT	Letters formed of straight lines or middle letter I or E	FIW, KEV, ZEH, YIK
CROSS	Letters can be re- arranged to spell word	HES, XOP, YOJ, MUH

TABLE 2  
MEAN NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES OF EACH OF FOUR CONCEPTS WITHIN EACH OF FOUR GROUPS

Group	Concept				
	DASH	TRI-ANGLE	DOT	CROSS	Total
3	4.50	1.69	1.37	2.19	9.81
6	4.31	2.56	1.44	1.75	10.06
9	4.81	2.25	2.37	2.12	11.56
18	4.12	3.36	1.56	1.81	10.81
Total of means	17.75	9.81	6.81	7.87	

TABLE 3  
ANALYSIS OF VARIANCE FOR REPEATED MEASURES ON TOTAL CORRECT IDENTIFICATIONS OF TEST INSTANCES

Source	df	SS	MS	F
Between subjects	63	120.44		
Between instances	3	13.50	4.50	2.52
Error (between)	60	106.94	1.78	
Within subjects	192	577.50		
Between concepts	3	204.03	68.01	35.79**
Interaction	9	30.78	3.33	1.75
Error (within)	180	342.69	1.90	

\*\*  $p < .01$ .

look up, and at 10 seconds were told the correct answer for the instance presented. The instance was exposed for 3 seconds following the reading of the correct answer. In the use of a correction procedure this experiment follows Callantine and Warren (1955).

After the presentation of the first 36 instances and again at the end of the presentation of the entire series of 72 instances, a blank card was exposed and subjects were asked to try to write the definition of each of the concepts.

Following the presentation of the series of 72 training instances of the concepts and the attempt to write the definitions of the concepts, the subjects were shown 20 additional new instances, five for each concept, and were told, "This time you will not be told the correct names of the examples. Remember, there is no penalty for guessing." The scores for each subject were the number of correct responses to each of the five test instances of the four concepts.

### Results

In Table 2 are presented means for the scores based on responses to the test instances. In all groups and for all concepts, identification scores exceed chance, albeit by very little in some cases. Analysis of variance (Table 3) indicated that only the difference in difficulty of the four concepts was significant. However, it became apparent that the four groups differed not only in the number of available positive instances of each concept, but in the

number of negative instances as well. Since a positive instance of DASH is also a negative instance of the other three concepts, the four groups were quite unequal with respect to negative instances. The possibility exists that an increase in negative as well as positive instances could increase cognitive strain to such an extent that information available from positive instances would be lost. A second experiment was planned to equate groups for negative instances while varying positive instances.

## EXPERIMENT II

### Procedure

Materials for concepts DASH, TRIANGLE, and CROSS were employed to construct new lists for three experimental conditions. In Condition 3 the subjects were exposed to three positive instances of one concept, each repeated six times, and six instances of each of the other two concepts, each instance being repeated three times. For Condition 9 subjects were exposed to nine instances of one concept, repeated twice, and six of each of the others. And, in Condition 18, subjects saw 18 instances of one concept, each only once, and again six of each of the others with three repetitions. All lists were randomized save for restrictions against repetition of an instance within a brief series.

Since the design was factorial with respect to number of instances of the concept, 3, 9, or 18, and concept difficulty, DASH, TRIANGLE, and CROSS, there were nine subgroups, each with seven subjects. The experimental procedures were much like those of Experiment I. Following presentation of all 54 instances, subjects were asked for written definitions of the concepts and were then tested with a series of eight new instances of each concept. The data for analysis are the number of correct responses for the single experimental concept during the test series.

### Results

The means for number of correct identifications of the experimentally manipulated concept in each of the nine subgroups are presented in Table

4. Again the means indicate that performance was better than chance in all groups—by a somewhat greater margin than in Experiment I.

Analysis of variance which is summarized in Table 5 yielded once more a highly significant effect attributable to differences between the concepts although, as reference to Table 4 will show, the principal difference is obviously the lower difficulty of DASH. As Table 4 also indicates TRIANGLE and CROSS are reversed in difficulty from Experiment I.

Of greater interest is the significant interaction between number of instances and concept. Inspection of the means would seem to indicate that for some reason nine instances, each repeated twice are less effective for DASH,

TABLE 4  
MEAN NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES OF THREE CONCEPTS IN EXPERIMENT II

Condition	Concept			
	DASH	TRIANGLE	CROSS	Total
3	7.86	3.57	6.27	17.70
9	6.14	6.57	4.57	17.28
18	8.00	4.29	4.71	17.00
Total	22.00	14.43	15.55	

TABLE 5  
ANALYSIS OF VARIANCE OF THE NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES IN EXPERIMENT II

Source	df	SS	MS	F
Number of instances	2	0.60	0.30	0.08
Concepts	2	77.74	38.87	10.62**
Instances $\times$ Concepts	4	61.41	15.35	4.19*
Error	54	198.06	3.66	
Total	62	337.81		

\*  $p < .05$ .

\*\*  $p < .01$ .

the very easy concept, than either fewer or more instances, that the very opposite is true for TRIANGLE, and that for CROSS a few instances repeated several times are better. The results from Experiment I would support the contention that TRIANGLE is not easily solved with very few repeated instances.

### DISCUSSION

Obviously the answer to the question of optimum number and frequency of instances to be presented in a concept attainment task is still obscure. The single interaction found pointed to the probable complexity of the answer, which will very likely be shown to be a function of rather specific characteristics of the concepts involved.

In Experiment II, for example, the three concepts differed, not only in difficulty, but in the characteristics of the instances on which they were based. DASH is based on the rather obvious characteristic of nonsense syllables, the particular letters from which they are formed. TRIANGLE is based upon characteristics of the letters themselves, and CROSS requires a mental manipulation or restructuring of the stimuli presented. It may well require the repetition of an instance on several occasions in order for most subjects to eliminate the alternatives more obvious than syllables with straight letters or rearranged words.

One difficulty which still existed in the procedure employed was that the contexts were not the same for all concepts. Thus, when TRIANGLE was being experimentally treated, it was being presented in the context of a difficult and an easy concept. DASH, by contrast, was always presented in the context of two more difficult concepts. Future work should make use of a common context of instances for all experimentally treated concepts.

It is believed that the discrepancy between our results and those of Callantine and Warren (1955) may result both from the different materials used and the fact that our experiments were adapted to group procedures. Further experiments are in progress which should throw some light on questions raised above.

### SUMMARY

Two experiments were carried out in an attempt to determine whether concepts are more easily discovered from the repeated examinations of a few instances of the concept or from single examinations of larger numbers of instances. Additionally, an attempt was made to vary the difficulty level of the concepts studied. The materials used consisted of concepts based upon non-sense syllables. In the first experiment there were no significant effects attributable to the number of different instances presented, but the conditions were thought to be confounded with varying numbers of negative instances. A second experiment was performed in which number of instances and type of concept were varied in a factorial design with number of negative instances being constant across conditions. Employing recognition of test instances as a measure, a significant interaction between number of instances and specific concept was found. Although not unequivocally interpretable the results appeared to indicate that when a concept is based upon less obvious characteristics of instances or required some mental transformation of stimuli, repetition of instances is desirable. Repetition may be less necessary with relatively obvious concepts.

### REFERENCES

- ADAMS, J. A. Multiple versus single problem training in human problem solving. *J. exp. Psychol.*, 1954, 48, 15-18.
- CALLANTINE, MARY F., & WARREN, J. M.

- Learning sets in human concept formation. *Psychol. Rep.*, 1955, **1**, 363-367.
- HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517-567.
- LUBORSKY, L. F. Aircraft recognition: I. The relative efficiency of teaching procedures. *J. appl. Psychol.*, 1945, **29**, 385-398.
- MORRISETT, L., JR., & HOVLAND, C. I. A comparison of three varieties of training in human problem solving. *J. exp. Psychol.*, 1959, **58**, 52-55.
- REED, H. B. Factors influencing the learning and retention of concepts: III. The origin of concepts. *J. exp. Psychol.*, 1946, **36**, 166-179.

(Received May 31, 1961)

# THE ENVIRONMENTAL ASSESSMENT TECHNIQUE: A WAY TO MEASURE COLLEGE ENVIRONMENTS<sup>1</sup>

ALEXANDER W. ASTIN AND JOHN L. HOLLAND

*National Merit Scholarship Corporation*

The purpose of this report is to describe a method for measuring the college environment, and to present some data on its validity and reliability. Even though it is axiomatic that, to predict behavior efficiently, we must assess the environment as well as the person, few techniques are available for assessing human environments. Barker and Wright (1954) have described an elaborate procedure for recording the child's experiences in his community. For assessing a more restricted type of community—the college environment or press—Pace and Stern (1958) have developed the College Characteristics Index (CCI), a true-false inventory by which students can describe the characteristics of the student body, faculty, and administration of the college. More recently, Lazarsfeld (1959) and Dailey (1960) have made suggestions for research which may produce other methods to assess the environment.

## RATIONALE

The Environmental Assessment Technique (EAT) is based on the notion, suggested by Linton (1945) and others, that a major portion of environmental forces is transmitted through other people. We can infer from this that the character of a social environment is dependent upon the nature of its members. Moreover, the dominant features of an environment are dependent upon the typical characteristics of its members. If, then, we

know the character of the people in a group, we should know the climate that group creates.

As a first step, we have assumed that the college environment or "press" is a product of the following attributes of the student body: the total number of students in the college, the average intelligence of the students, and the personal characteristics of the student body (as estimated by a typology of six types). These eight variables are assumed to encompass an extensive range of the student's attributes, including his personality, interests, values, originality, self-concept, parental background, goals, and aptitudes. Assessing colleges by means of the EAT should yield environmental descriptions which will enable us to study the interactions of students and colleges in terms of a variety of educational and personal outcomes.

## ENVIRONMENTAL MEASURES

Explicit definitions of the eight EAT variables are given in the following sections.

*Institutional Size.* The size of the student body ( $N$ ) is simply the total enrollment at the institution. Since the distribution of college enrollments tends to have a markedly positive skew, the square root of  $N$  was chosen as the actual measure of size. This transformation simplifies computations by reducing the number of digits and makes the distribution of college size more nearly normal.

*Intelligence Level of the Student Body.* To secure intelligence estimates of the student bodies, mean scores on the National Merit Scholarship Qualifying Test (NMSQT) were obtained for a sample of students entering undergraduate institutions. These students came from a 10% national sample of high school juniors scoring above the sixty-fifth percentile on the 1959 NMSQT. Only the scores of students

<sup>1</sup> This study is a part of the research program of the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation.

who actually entered college were used. Colleges attended by less than 15 of these students were excluded, leaving a total of 335 four-year, degree granting institutions.

The following check was made on the representativeness of differences between mean intelligence levels in the smaller samples ( $N$ s between 15 and 24): mean scores on the CEEB Scholastic Aptitude Test, which were available for the entire 1956 freshman class (Fishman, 1957) at 37 of the 165 institutions enrolling less than 25 NMSQT students, were correlated with the corresponding mean NMSQT scores. The Pearson coefficient of .92 indicated satisfactory representativeness, particularly in view of the fact that different tests, given in different years, were being compared.

*Personal Orientation of the Student Body.* The method of personality assessment used in EAT is based on Holland's (1959b, 1961) theory of vocational choice. A central hypothesis of this theory is that persons in different occupations have different personalities. Holland has proposed a classification of occupations in terms of six personality types: Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic. For each of these types there is a narrative summary of personal characteristics called the "model personal orientation." By identifying the type to which any vocation belongs, we can use a person's vocational choice as a miniature personality "test." For example, if a person's choice is engineering, which falls in the Realistic class, we would expect him to possess some of the characteristics of the model Realistic orientation: masculine, physically strong, unsociable, aggressive, etc. The validities of these personal orientations have been examined in several recent studies (Holland, 1959a, 1960, 1961).

Using this theory as a beginning, we have assumed that a given social environment can, to some degree, be described in terms of the occupations (personalities) of its members. To take a simple example: We would expect the environment or "climate" of a law firm to differ from that of an engineering firm, to the extent that we knew engineers and lawyers to be different kinds of people. If we extend this notion to the college, it should be possible to characterize the college environment in terms of the college environments (occupations) of the students.

As a first step toward characterizing the environments of the 335 institutions, each college major field was classified as belonging in one of Holland's six classes. The assignment of major fields to classes was made by using as guides the original formulations

in the theory and the keys (six lists of occupational titles) in the Holland Vocational Preference Inventory (Holland, 1959a). Cases of doubt were resolved in discussion by the authors.

Descriptions of the model personal orientations and lists of the major fields included in each orientation are given in Table 1. These descriptions represent hypotheses about the climate and "press characteristics" associated with each type. Information about the number of students at a given institution who obtained a baccalaureate degree in a specific major field was secured from the publication, *Earned Degrees Conferred by Higher Educational Institutions: 1957-1958* (United States Office of Education, 1959). Unclassifiable degrees, representing about 3.2% of the total, were omitted from our analysis.

To obtain measures of the six personal orientations at any institution, the number of majors (i.e., students in a major field) of each orientation was expressed as a percentage of the total number of classifiable majors. For instance, the Realistic orientation of a college environment would be:

$$\text{Realistic orientation} =$$

$$\frac{\text{Number of majors classified Realistic} \times 100}{\text{Total number of classifiable majors}}$$

In this way, each institution was characterized by the proportion of its students belonging to each of the six personal orientations, as determined by their choice of major field.

#### VALIDITY OF THE EAT

An initial test of the EAT's validity was made using the College Characteristics Index. In the CCI, observers (usually students) at the institution describe the college in terms of 300 dichotomous items. These items are scored on 30 press scales of 10 items each. Some of these CCI data, which were obtained in 1958 from Merit Scholars and Finalists at 36 institutions,<sup>2</sup> have been used in two previous

<sup>2</sup> Amherst, Brown, California Institute of Technology, Carleton, Carnegie, University of Chicago, Colorado, Cornell, Dartmouth, Duke, Georgia Institute of Technology, Harvard, Indiana, Iowa State, Kansas, Massachusetts Institute of Technology,

TABLE 1  
COLLEGE MAJOR FIELDS CORRESPONDING TO EACH OF SIX PERSONAL ORIENTATIONS

Orientation	Description (from Holland, 1961)	Relevant major fields
Realistic	"masculine, physically strong, unsociable, aggressive . . . prefers concrete to abstract"	agriculture, agricultural education, physical education, recreation, industrial arts, engineering, forestry, trade, and industry
Intellectual	"task-oriented, intraceptive, asocial, prefers to think through rather than act out; needs to understand"	architecture, biological sciences, geography, medical technology, pharmacy, mathematics, philosophy, physical sciences, anthropology
Social	"sociable, responsible, feminine . . . needs attention . . . avoids intellectual problem-solving . . . orally dependent"	health education, education of exceptional children and mentally retarded, speech correction, education (unclass.), nursing, occupational therapy, physical therapy, scholastic philosophy, social science (general), American civilization, sociology, social work
Conventional	"prefers structured numerical and verbal activities and subordinate roles . . . conforming . . . identifies with power, externals, and status"	accounting, secretarial, business and commercial (general and unclass.), business education, library science, economics
Enterprising	"verbal skills for dominating, selling, leading others . . . orally aggressive"	hotel and restaurant administration, hospital administration, history, international relations, political science, foreign service, industrial relations, public administration
Artistic	"asocial; avoids problems which are highly structured or require gross physical skills . . . intraceptive . . . need for individualistic expression"	art education, music education, English and journalism, fine and applied arts (all fields), foreign language and literature (all fields)

studies (Astin, 1961; Thistlethwaite, 1959). Each of the eight EAT measures ( $\sqrt{N}$ , Intelligence, and the six personal orientations) was correlated with the 30 CCI scales (Table 2). Of the 240 coefficients in Table 2, 23% are significant at the .01 level, and 39% at the .05 level.

#### The CCI scales most highly related

Michigan, Minnesota, Northwestern, Notre Dame, Oberlin, Pennsylvania, Pomona, Princeton, Purdue, Radcliffe, Rensselaer Polytechnic Institute, Rice, Smith, Stanford, Swarthmore, University of Texas, Wisconsin, Wellesley, Wesleyan, and Yale.

to institutional size are Aggression (.64), Achievement (-.59), Counteraction-Infavoidance (-.58), Understanding (-.58), Passivity (.55), Fantasied Achievement (-.55), Deference (.54), Sex (.54), Exhibition (.53), and Pragmatism (.52). These results suggest that in the larger as opposed to the smaller student bodies there is more aggressive behavior, exhibitionism, and heterosexual activity, and more deference shown toward faculty. In smaller student bodies, on the other hand, there is more academic competi-

TABLE 2  
THE RELATION OF EIGHT ENVIRONMENTAL MEASURES TO THE  
COLLEGE CHARACTERISTICS INDEX AT 36 INSTITUTIONS

CCI scales	Size of student body	Intelli-gence of student body	Personal orientations <sup>a</sup>					
			Real-istic	Intel-lectual	Social	Conven-tional	Enter-prising	Artistic
Abasement	20	-37	38	-29	.04	11	-24	-45
Achievement	-59	64	-20	33	-22	-35	42	35
Adaptiveness	14	-55	14	-39	26	13	-25	-20
Affiliation-Rejection	-35	.08	-21	38	-01	-07	.05	16
Aggression-Blamavoidance	64	.09	45	.09	-25	.08	-38	-58
Change-Sameness	-06	17	20	14	-14	-02	-10	-09
Conjunctivity-Disjunctivity	-19	.06	-06	16	-07	-24	22	12
Counteraction-Infavoidance	-58	49	.06	28	-41	-29	11	19
Deference	54	-63	26	-55	20	30	-38	-26
Dominance	-19	21	-21	32	.08	-02	.03	.09
Ego Achievement	-02	-19	-34	-13	38	12	15	24
Emotionality-Placidity	-33	23	-25	21	-14	18	20	14
Energy-Passivity	-55	48	-05	31	-35	-42	28	33
Exhibition-Infavoidance	53	-45	-14	-33	43	36	-15	-03
Fantasied Achievement	-55	49	-32	46	-17	-21	35	38
Harmavoidance	-30	26	-50	.04	21	-14	55	56
Humanism	-28	55	-81	23	25	.07	79	64
Impulsion-Deliberation	27	.00	11	28	-08	23	-28	-40
Narcissism	49	-36	-17	-43	59	.01	.01	18
Nurturance-Rejection	-03	-26	-37	-18	39	-06	22	46
Objectivity	-42	68	-26	46	-24	-26	40	31
Order	28	-51	-05	-31	36	15	-10	-03
Play	47	-60	25	-34	35	.04	-36	-30
Pragmatism	52	-67	73	-32	-08	.02	-73	-66
Reflectiveness	-29	50	-62	16	14	.01	64	54
Scientism	-18	31	33	38	-43	-08	-22	-48
Sentience	-28	47	-70	10	31	-13	71	69
Sex-Prudery	54	-40	-22	-24	53	15	.01	.06
Succorance-Autonomy	-01	-28	-09	10	14	24	-14	-13
Understanding	-58	70	-28	46	-23	-24	41	33

Note.— $r_{05} = .33$ ,  $r_{01} = .43$ . Decimal points omitted in table.

<sup>a</sup> The percentage of college majors falling in each class.

tion, more striving for intellectual achievement, and greater involvement in campus activities (Counteraction-Infavoidance, Passivity).

Intelligence level of the student body had 18 significant correlations with the CCI scales, more than any of the other seven EAT variables. The highest correlates of intelligence were Understanding (.70), Objectivity (.68), Pragmatism (-.67), Achievement (.68), Deference (-.63), Playment (.64), Adaptiveness (-.55), Humanism (-.60), Adaptiveness (-.55), Order (-.51), and Reflectiveness (.50).

Most of these press characteristics are those we would expect to find in the more intelligent student bodies: an emphasis on achievement, intellectual objectivity, seriousness of purpose, and a need to analyze and understand, rather than to order and structure.

The CCI variables most closely related to the first personal orientation, Realistic, were Humanism (-.81), Pragmatism (.73), Sentience (-.70), Reflectiveness (-.62), and Harmavoidance (-.50). These correlations

imply a preference for the practical, the concrete rather than the abstract, and an aversion to feeling experience—traits which are consistent with the descriptions of the Realistic orientation (Table 1) developed earlier (Holland, 1959b).

The Intellectual orientation had its highest correlations with Deference (-.55), Fantasied Achievement (.46), Objectivity (.46), and Understanding (.46). These results imply that college environments with high Intellectual orientations have many of the expected characteristics: independence of thought and action, great need to achieve, and a need to analyze and understand.

The CCI scales correlating highest with the Social orientation were Narcissism (.59), Sex (.53), Exhibition-Infavoidance (.43), and Scientism (-.43). Here there is a close correspondence with the hypothetical description (Holland, 1961): "sociable, responsible, feminine...needs attention...avoids intellectual problem-solving...orally dependent" (Table 1).

The three highest correlates of the Conventional orientation, Passivity (.42), Exhibition-Infavoidance (.36), and Achievement (-.35), while of only borderline significance, are also generally consistent with the description given in Table 1.

The Enterprising and Artistic orientations are similar in the patterns of their relationships with the CCI scales: Humanism (.79 with Enterprising and .64 with Artistic), Pragmatism (-.73 and -.66), Sentience (.71 and .69), Reflectiveness (.64 and .54), Harmavoidance (.55 and .56), and Blamavoidance (.38 and .58). Perhaps one of the reasons that these two orientations turned out to be so similar, aside from the fact that they are known to be related (Holland, 1959a, p. 29) is that many of the stu-

dents majoring in "business and commercial," all of whom were included in the Conventional orientation, might more properly be included in the Enterprising orientation, particularly those who plan careers in sales, promotional, and related areas.

The characteristics of the Enterprising and Artistic orientations are just the opposite of those associated with the Realistic orientation. While this is again consistent with previous descriptions (Holland, 1959a), there is the suggestion that these three orientations may be somewhat redundant.

Another partial test of EAT's validity was attempted using data reported in the book, *What College Students Think* (Goldsen, Rosenberg, Williams, & Suchman, 1960). These investigators attempted to assess and compare the values, attitudes, and opinions of students at the following institutions: Cornell, Dartmouth, Fisk, Harvard, Michigan, North Carolina, Texas, UCLA, Wayne, Wesleyan, and Yale. Even though the number of institutions was quite small, it seemed likely that the environments of these colleges would differ markedly on the variable of Intelligence. Accordingly, 16 of the questions asked of the student bodies were selected on the basis of their hypothesized relation to Intelligence. These questions dealt with: the amount of cheating admitted to by the students; preference for Republican as opposed to Democratic party; belief in God; importance of religion; attitudes toward labor laws, the welfare state, minimum wage, and a free college education; importance in college of learning values; importance of athletics; belief in the existence of academic freedom; "tendency to personify the fraternity"; and the relative importance of several possible goals of a college education—general education, vocational training, training in how to

get along with people, preparation for a happy marriage.

The institutions were ranked (1-11) on Intelligence using the sample mean NMSQT scores (*Ns* ranged from 15 to 203); they were also ranked (1-11) on each of the 16 questions on the basis of the percentage of students giving positive responses, so that rank-difference correlations could be computed. Seven of these rho's were significant at the .05 level. The relevant variables and their correlations with Intelligence were:

A college ought to emphasize preparation "for a happy marriage and family life."	-.99
"College teachers are afraid to say what they really believe these days."	-.84
A college ought to "provide vocational training, develop skills and techniques directly applicable to your career."	-.81
Religion is a "major source of satisfaction."	-.80
A college ought to emphasize "a basic general education and appreciation of ideas."	.75
Republican political party preference	.73
Belief in God	-.69

On an exploratory basis, 14 more rank correlations were computed between the other seven EAT measures and two of the items which had not correlated significantly with Intelligence. Two more significant rho's were obtained: the Conventional orientation was positively related (.73) to the "tendency to personify the fraternity" and frequency of admitting cheating was negatively related (-.70) to the Enterprising orientation.

To illustrate in more concrete terms the characterization of colleges via the EAT, the three institutions (from the original sample of 335) scoring highest on each orientation were identified (Table 3). The institutions with highest Realistic orientations are technologically institutions, primarily or ex-

TABLE 3  
HIGHEST SCORING OF 335 INSTITUTIONS  
ON EACH PERSONAL ORIENTATION

Realistic	Worcester Polytechnic Institute Michigan College of Mining and Technology Case Institute of Technology
Intellectual	Union College and University California Institute of Technology Reed College
Social	State Teachers College at Towson (Maryland) Danbury State College (Connecticut) Georgia State College for Women
Conventional	Seton Hall University Xavier University (Ohio) University of Scranton
Enterprising	Georgetown University John Carroll University Wabash College
Artistic	College of New Rochelle Emmanuel College (New York) Hollins College

clusively for men; in contrast, those highest on the Artistic orientation are small liberal arts colleges for women. The highest scorers on the Social orientation include two coeducational teachers' colleges and a liberal arts college for women. Those scoring highest on the Conventional orientation are all large universities controlled by the Roman Catholic Church. (In fact, seven of the eight institutions ranking highest on Conventional are Roman Catholic institutions.)

All three of the institutions high on the Intellectual orientation are known for their strong emphasis on scholarship and research. The California Institute of Technology, despite its name, is classified as a "college, gradu-

TABLE 4

TEST-RETEST RELIABILITIES OF THE SIX ENVIRONMENTAL PERSONAL ORIENTATIONS FOR INTERVALS OF 1, 3, AND 6 YEARS (31 INSTITUTIONS)

Orientation	Reliability		
	1 Year (1955-56)	3 Years (1955-58)	6 Years (1952-58)
Realistic	.99	.98	.97
Intellectual	.81	.55	.54
Social	.88	.87	.80
Conventional	.94	.93	.93
Enterprising	.96	.84	.83
Artistic	.95	.96	.93

ate school, and institute of research in science, engineering, and the humanities" (Irwin, 1960, p. 186).

Two of the three highest scorers on Enterprising, Georgetown University and Wabash College, are of particular interest in view of the hypothesized correlates of that orientation: "verbal skills for dominating, selling, leading others... orally aggressive." Georgetown University is noted for its large Law School and its School of Foreign Service. Wabash College, too, with its Personal Development Program for "young executives," its Institute of Politics, and its Washington Semester Program for "students of demonstrated ability and interest in public affairs" (Irwin, 1960, pp. 391-392), appears to place considerable emphasis on "enterprising" activities.

#### RELIABILITY OF THE PERSONAL ORIENTATIONS

A check on the test-retest reliability of the six measures of personal orientation was made using a 10% random sample ( $N = 31$ ) of institutions from the original sample of 335.<sup>3</sup> Scores on

<sup>3</sup>The entire sample of colleges was not used because of the prohibitive amount of clerical labor involved in collating major fields from *Earned Degrees Conferred by Higher Educational Institutions: 1957-1958* (United States Office of Education, 1959) by college.

the six personal orientations for these 31 colleges were obtained for the graduation years of 1952, 1955, 1956, and 1958 by collating the major fields listed in *Earned Degrees Conferred by Higher Educational Institutions: 1957-1958* (United States Office of Education, 1959) for those years. Table 4 shows retest reliabilities over intervals of 1, 3, and 6 years. With the possible exception of the Intellectual orientation, all measures appear to be highly reliable over the 1-year interval. Even over the 6-year period, all orientations except the Intellectual seem to have satisfactorily high reliabilities. (Table 4 also shows that differences among institutions in their curricula are highly stable from year to year.)

#### INTERCORRELATIONS OF EAT VARIABLES

Intercorrelations of the eight EAT variables are shown in Table 5. While all but four of the 28 correlations are significant ( $p < .05$ ), the highest correlation is only  $-.57$  (between the Realistic and Artistic orientations). An inspection of the larger correlations reveals two major clusters: the first combines Intelligence, the Intellectual orientation, and the Enterprising orientation; the second, involving the Realistic (negatively), Enterprising, and Artistic orientations, was to be expected from the patterns of CCI correlations (Table 2) discussed previously.

It should be noted that the six personal orientations are self-limiting (i.e., for any college the six scores must sum to 100). Such dependence among these measures accounts at least partially for the fact that 12 of the 15 correlations among the personal orientations were in a negative direction.

#### DISCUSSION

The results imply that this new method (EAT) for assessing college

TABLE 5  
INTERCORRELATIONS OF EAT VARIABLES  
(N = 335)

	2	3	4	5	6	7	8
1. Size	-.09	.26	-.28	-.06	.26	-.25	-.29
2. Intelligence		-.08	.41	-.36	-.27	.48	.34
3. Realistic orientation			-.24	-.37	-.17	-.46	-.57
4. Intellectual orientation				-.40	-.17	.38	.16
5. Social orientation					-.23	-.32	-.01
6. Conventional orientation						-.12	-.30
7. Enterprising orientation							.45
8. Artistic orientation							

Note.— $r_{08} = .11$ ;  $r_{01} = .14$ .

environments possesses moderate validity and substantial reliability. Moreover, EAT, being limited to eight variables which are inexpensive to obtain,<sup>4</sup> lessens the computational problems which arise in the study of college environments and brings such studies within the financial means of more researchers.

One limitation of the present study is that it used the major fields of the students, rather than their vocational choices. Since some of the major fields reported in *Earned Degrees* are ambiguous with regard to the actual occupations implied, it was necessary to make difficult and perhaps arbitrary decisions in assigning them to a particular class. As was noted previously, this is particularly true of those "business" majors who plan careers in sales or advertising, and should be classified as Enterprising instead of Conventional. Some of these problems would be solved if we knew the student's vocational choice.

In addition, the present scheme does not give direct information about the nature of the faculty. It is likely, however, that in level of ability, number,

<sup>4</sup> Intelligence level of the student body could, of course, be expensive to obtain; however, most institutions routinely assess their incoming freshman classes with standardized aptitude or achievement tests, and even in the absence of such test data the average high school rank of the students would probably give a useful estimate of intelligence level in the student body.

and "primary occupations" (fields of specialization), the faculty of an institution resembles its student body. This assumption is supported by the fact that many of the CCI items are actually descriptions of specific faculty behaviors. In addition, Pace and Stern (1958) found that faculty and student reports of the college press were similar.

EAT appears to be especially well-suited to the study of college-student interactions. A major problem in such studies is how to assess the similarity between persons and environments in a meaningful way. Since the EAT is based principally on a heuristic theory of vocational choice and occupational types, it permits us to classify both students and colleges according to the same six types. More importantly, the same operations (vocational choice or major field) are used for typing both the student and his environment.

The present results also indicate that the attributes of the student body reflect a major portion of what has been called the college press or environment,<sup>5</sup> since the press characteris-

<sup>5</sup> It should be pointed out that such a conclusion is not necessarily at variance with studies such as those of McFee (1961), who found no evidence at one college of a systematic relationship between the needs of individual students and their perceptions of the college press. The question of the environment's being related to the nature of the students concerns aggregates of student characteristics across institutions.

ties (as measured by the CCI) of a particular school can be predicted with modest success simply from a knowledge of the students' number, average intelligence, and distribution of major fields. This conclusion is supported by a recent study of college productivity (Astin, 1961), which showed that, even for small samples of students *entering* college, the proportion planning to major in natural science was highly related to several student and faculty press variables. It is obvious that in future studies of college "influence," the students' characteristics must be taken into account before we attribute differences in performance to the college press.

One of the most compelling aspects of EAT is the possibility it offers for measuring other kinds of environments. If, as the present data suggest, a simple occupational census of the members of any group gives a valid estimate of the environment or climate of that group, it should be possible to conduct a wide variety of cultural and sociological studies using an elementary method. Studies of other institutions and cross-cultural studies, in particular, would seem especially suited to this method.

### SUMMARY

A method for measuring the college environment, the Environmental Assessment Technique (EAT), was described. Measures of eight characteristics of the student body—its size, average intelligence, and six "personal orientations"—were validated against the College Characteristics Index at 36 colleges. The six personal orientations, which were obtained from a census of students' major fields at each institution, were also shown to be highly stable over time. Some of the advantages, limitations, and possible applications of the EAT were discussed.

### REFERENCES

- ASTIN, A. W. A re-examination of college productivity. *J. educ. Psychol.*, 1961, **52**, 173-178.
- BARKER, R. G., & WRIGHT, H. F. *Midwest and its children*. Evanston, Ill.: Row, Peterson, 1954.
- DAILEY, C. A. Natural history and phenomenology. *J. indiv. Psychol.*, 1960, **16**, 36-44.
- FISHMAN, J. A. *1957 supplement to College Board scores no. 2*. New York: College Entrance Examination Board, 1957.
- GOLDSEN, ROSE K., ROSENBERG, M., WILLIAMS, R. M., JR., & SUCHMAN, E. A. *What college students think*. Princeton: Van Nostrand, 1960.
- HOLLAND, J. L. *Preliminary manual for the Holland Vocational Preference Inventory*. (Res. ed.) Palo Alto, Calif.: Consulting Psychologists Press, 1959. (a)
- HOLLAND, J. L. A theory of vocational choice. *J. counsel. Psychol.*, 1959, **6**, 35-45. (b)
- HOLLAND, J. L. The relation of the Vocational Preference Inventory to the Sixteen Personality Factor Questionnaire. *J. appl. Psychol.*, 1960, **44**, 291-296.
- HOLLAND, J. L. Some explorations of a theory of vocational choice. Evanston, Ill.: National Merit Scholarship Corporation, 1961. (Mimeo)
- IRWIN, MARY (Ed.) *American universities and colleges*. (8th ed.) Washington, D. C.: American Council on Education, 1960.
- LAZARSFELD, P. F. Progress report on a project to map out the general area of non-intellectual factors in the prediction of college success. New York: College Entrance Examination Board, 1959. (Mimeo)
- LINTON, R. *The cultural background of personality*. New York: Century, 1945.
- McFEE, ANNE. The relation of students' needs to their perception of the college environment. *J. educ. Psychol.*, 1961, **52**, 25-29.
- PACE, C. R., & STERN, G. G. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269-277.
- THISTLETHWAITE, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183-191.
- UNITED STATES OFFICE OF EDUCATION. *Earned degrees conferred by higher educational institutions: 1957-1958*. (Circular No. 570) Washington, D. C.: United States Department of Health, Education, and Welfare, 1959.

(Received June 4, 1961)

## THE DESIRE FOR SECURITY: AN ELEMENT IN THE VOCATIONAL CHOICE OF COLLEGE MEN<sup>1</sup>

STUART H. BLUM<sup>2</sup>

*Teachers College, Columbia University*

Considerable investigation has been carried out by workers of many disciplines in attempting to determine the manner in which various factors are related to vocational choice. Factors believed to be operating in this choice are numerous and include such concepts as overall mental capacity, specific aptitudes, interests, and temperament. One of the often mentioned, but less investigated, of these factors is the desire for security.

Interviewing and counseling experience with both military personnel and civilian college students on the part of the investigator has yielded the impression that a desire for security is one of the most conspicuous factors involved in the selection of an occupation or in the choice of a specific job. Expressed emphasis by interviewees on such factors as fringe benefits, tenure, guarantee of employment, and on the use of the term security itself, often far outweighed concern about the content or nature of the actual work. Increasing trends toward a so-called Welfare State and both contemporary nonfiction and reports of scientific studies suggest an increased interest on the part of the general population in obtaining security.

<sup>1</sup>The article is based upon a PhD dissertation submitted to Teachers College, Columbia University. The author wishes to express his appreciation for the guidance and encouragement received from the members of his Dissertation Committee, Robert L. Thorndike (Chairman), Elizabeth P. L. Hagen, and the late Irving D. Lorge.

<sup>2</sup>Now at the Testing and Counseling Center, Hofstra College, Hempstead, New York.

The purpose of the study on which this report is based was to ascertain differences among individuals with respect to the degree of importance placed upon a desire for security as a factor in vocational choice, and to determine how this degree of emphasis is related to other pertinent characteristics or variables of the individual. Whereas any measure of security or insecurity is probably a reflection of a deeper psychological security, for purposes of the present investigation the deeper psychological aspects were largely treated as measurable normal personality variables.

The areas for investigation were based on the expectation that:

1. Degree of emphasis placed upon security as a factor in vocational choice exists as a trait of some comprehensiveness that can be measured reliably.

2. A positive relationship exists between emphasis upon security in vocational choice and the actual choice of a job situation of a secure type.

3. Significant relationships exist between emphasis upon security in vocational choice and normal personality variables as measured by the Edwards Personal Preference Schedule (1957).

4. Significant relationships exist between emphasis upon security in vocational choice and background factors.

Since an instrument to measure the degree of emphasis placed upon security in vocational choice was not available that met desired specifications, a 40-item inventory was developed.

### *The Security Inventory<sup>3</sup>*

The final form of the Security Inventory contained 40 forced-choice items of the triad type. Of the three alternatives contained in each item one was scored as contributing to the Security variable which was thus expressed as a simple raw score, with a possible range from 0-40.

The content of the items was based on the investigator's interviewing and counseling experience, and utilized concepts expressed in Caplow (1954), Daugherty (1941), Hayek (1944), Hopcock (1935), Lindahl (1949), Lorge and Blau (1942), McGregor (1944), Nolan (1956), Riesman, Denny, and Glazer (1953), Roe (1956), Shartle (1952), Thorndike and Hagen (1957), Viteles (1953), and Whyte (1956). The basic conceptualization of a desire for security as expressed in the items included, when compared to other alternatives, a preference for: membership in a group; fringe and retirement benefits; freedom from the ups and downs of business or political conditions (including the concept of a guaranteed annual wage); dependence on rules, regulations, standards, procedures, and decisions of superiors (freedom from responsibility on the job); physical safety; allowance for personal shortcomings of the individual; existence of similar jobs in other organizations or under conditions of national emergency; lack of worry, risk, and competition; assurance of the maintenance of a certain standard of living; personal comfort; a fixed overall work

schedule; compensation, promotion, and retention, based on amount and length of time worked; working for an organization (rather than being own boss); the proven and the established; the familiar; routine and certainty.

The booklet of the Security Inventory was entitled, *Attitudes in Vocational Choice*, and the subjects were not informed that desire for security was being measured. All items were to be completed in spite of any difficulty of making a choice.

For inclusion in the Security Inventory, the selection of the final items from an original pool of 80 was primarily based on agreement by judges as to the Security response in each item and upon a subsequent item analysis with a preliminary group of 122 male college students.

### PROCEDURE

In addition to the Security Inventory and the Edwards Personal Preference Schedule, a 24-item biographical information questionnaire which included items designed to obtain some indication of the actual job situation desired by each individual as well as his characteristics and experiences considered pertinent to the study, was prepared.

The complete set of instruments was administered to 513 students enrolled in the following institutions: City College (New York), Fairleigh Dickinson University, Colgate University, Buffalo University, and Columbia University. Of this number 404 students met the qualifications of bona fide male upperclassmen (juniors and seniors aged 30 or less). In addition to practical considerations such as availability of subjects, ease of administration of instruments, and motivation for cooperation, the male college upperclassmen group provides information and suggests implications for both education and industry as well as for psychology. Further, this group represents the potential leadership in many areas of our society at a period in time when the group's members are focusing on their vocational choices. Because the study was concerned not with absolute distribution, but rather with relationships between results on various measures, the primary concern in the selection of subjects was to secure a heterogeneous sample of upperclassmen in order

<sup>3</sup>A copy of the complete Security Inventory as well as a detailed description of its construction may be found in the dissertation, *Security in Vocational Choice: A Study of Male College Upperclassmen*, available from University Microfilms, Incorporated; 313 North First Street; Ann Arbor, Michigan; Order No. Mic 60-5083 remitting \$3.35 for microfilm, or \$4.80 for xerographic enlargement.

that adequate range of response would be obtained.

Inasmuch as the Security Inventory yielded the variable against which all the other variables examined were to be compared, the limited class time was used for the administration of this instrument. The subjects completed the Edwards Personal Preference Schedule and the biographical information questionnaire out of class on their own. A system of individual code numbers provided anonymity, and as is consistent with obtaining some degree of motivation, and with fairness and ethical standards, there was some feedback of results. Of the 404 subjects who met the qualifications for inclusion in the sample, 86%, or 346 students, completed all the materials.

Raw scores (the total number of "security alternatives" chosen) on the Security Inventory and the raw scores on the Edwards Personal Preference Schedule were correlated. Wherever possible, the items of the biographical information questionnaire were coded so as to make possible product-moment correlations with the raw scores of the Security Inventory. Analysis of variance procedures were utilized where correlational analysis was not feasible.

### RESULTS

The successful development of the Security Inventory was obviously crucial to further pursuit of the investigation. On the basis of 404 male juniors and seniors, comprising the sample of the main study before eliminations due to deficiencies on other instruments, a mean of 13.6 and a standard deviation of 6.0 was obtained for the final form of the inventory. The scores had a basically unbroken and symmetrical distribution. The reliability as estimated by Kuder-Richardson Formula Number 20 was found to be .82.

Table 1 shows the results of the analysis of variance used to test the significance of the differences among the means of the security scores for each of the categories of anticipated job situation. Progressively higher security score means are associated with progressively larger and generally more secure employers. The large size of the  $F$  ratio obtained indicates dif-

TABLE 1  
SECURITY SCORE AND INDICATION OF  
ANTICIPATED JOB SITUATION  
( $N = 346$ )

Anticipated job situation	<i>M</i>	<i>SD</i>	<i>N</i>
Work for self	11.27	5.19	94
Work in civil service	17.94	5.80	35
Work for a large company	14.44	5.97	176
Work for a small company	13.88	5.27	41

$$F = 13.07, p < .01$$

ferences significant at the .01 level. It should be noted that the mean *security* score of those who anticipate working for themselves is approximately a full standard deviation below the mean security score of those who plan to work in civil service. The above findings also serve to validate further the Security Inventory itself, for the scores on the inventory are in the direction of the actual job situation choices.

Table 2 presents the correlations of the security score with each of the personality variables of the EPPS. Eight of the 15 coefficients and the multiple correlation coefficient of security score with the 15 EPPS variables are significant at the .01 level. One additional coefficient is significant at the .05 level only. Each relationship is in a direction consistent with the basic conceptualization of a desire for security. The personality characteristics most closely associated with emphasis on security in choosing a job or an occupation, are a desire for Order and an avoidance of Change.

Of the biographical variables examined by correlational analysis, two yielded coefficients with security score significant at the .01 level. High desire for security in choosing a job or an occupation shows a slight tendency to be associated with lower status of father's occupation ( $r = .18$ ) and with frequent religious service attendance

TABLE 2  
MEANS AND STANDARD DEVIATIONS OF THE EDWARDS PERSONALITY VARIABLES AND CORRELATION COEFFICIENTS, REGRESSION WEIGHTS, AND THE MULTIPLE CORRELATION COEFFICIENT OF THE SECURITY SCORE WITH THESE VARIABLES  
(*N* = 346)

Variable	<i>M</i>	<i>SD</i>	Coefficient	Regression weight
Security	13.87	5.99		
Achievement	16.92	4.23	-.22**	-.13
Deference	10.72	3.28	.17**	.07
Order	11.05	4.44	.32**	.22
Exhibition	14.69	3.53	-.05	.06
Autonomy	14.92	4.28	-.27**	-.11
Affiliation	13.50	3.87	.09	.08
Intraception	15.86	4.69	-.06	-.02
Succorance	9.76	4.37	.21**	.11
Dominance	17.36	4.72	-.22**	-.10
Abasement	11.59	4.74	.23**	.12
Nurturance	12.82	4.58	.12*	0
Change	15.52	4.57	-.34**	-.21
Endurance	13.90	5.12	.10	.05
Heterosexuality	17.42	5.23	.02	.10
Aggression	13.97	4.43	-.08	.02
Consistency	11.45	2.09	-.05	
Multiple			.54**	

Note.—Coefficients are all product-moment.

\* Significant at .05 level.

\*\* Significant at .01 level.

TABLE 3  
SECURITY SCORE AND FIELD OF MAJOR STUDY IN COLLEGE  
(*N* = 346)

Field of major study	<i>M</i>	<i>SD</i>	<i>N</i>
Business	14.94	5.90	118
Engineering	13.84	4.87	49
Humanities	13.00	7.16	21
Pre-law	11.13	7.95	16
Pre-medical and pre-dental	12.12	4.23	33
Science	12.75	5.34	51
Social Studies	14.36	6.89	42
Teaching	15.81	5.57	16

$$F = 2.06, p < .05$$

(*r* = .18). Three of the variables yielded coefficients with security score significant at the .05 level only. Emphasis upon security in choosing a job or an occupation shows a slight tendency to be associated with poor health

(*r* = .12), little athletic participation in college (*r* = .11), and a shorter response to the question concerning occupational goals (*r* = .11).

Of the biographical variables examined by analysis of variance, only the means of the security scores for each of the fields of major study in college gave an *F* ratio significant at the .05 level.

The hypothesized expectations were thus confirmed, and for college upper-classmen at least, it can be concluded that emphasis upon a desire for security in choosing a job or an occupation exists as a trait of some comprehensiveness that can be measured reliably, is positively related to the actual choice of a job situation of a secure type, and is characterized by low, but statistically significant personality and background correlates.

#### DISCUSSION

It is likely that what might be classed as chance or accident plays an important role in an individual's selection of a job or an occupation. Nonetheless, previous investigations have pointed to the relative importance of the desire for security as a factor in this decision. The investigation on which this report is based has demonstrated that although statistically significant relationships exist between the variable and both normal personality variables as measured by the EPPS and background factors, these relationships are small. For practical purposes, therefore, it would often seem desirable to include a measure of an individual's desire for security in choosing a job or an occupation as a part of vocational guidance and personnel selection procedure. The results of a more comprehensive and clinically derived measure of basic psychological security-insecurity might well possess a high positive relationship with the score obtained on the Security Inventory. However, from the view-

point of administrative efficiency, as well as from that of validity for the vocational guidance and/or personnel selection situation, the use of the more specific device appears indicated.

The differences among security score means for each of the fields of major study in college provide the basis for an additional measure in guidance practice. For example, a high security score would be deviant for a pre-law student, and might suggest further interview, since the group comprising this academic major possesses the lowest mean score.

It may be argued that the measure of desire for security is contaminated by the educational or occupational choice the person has already made. When an individual has chosen a field of major study in college that prepares for a specific profession or vocation, he has to some extent made an occupational commitment. He may have chosen the field for reasons other than a desire for security. Yet the choice may have influenced him to select alternatives that were consistent with this choice when filling out the Security Inventory. The individual who makes an occupational commitment, presumably on the basis of other than security reasons or motives, nonetheless, has probably in some manner taken security into account. It has not been determined which of these two elements of motivation takes precedence.

#### SUMMARY

One of the often emphasized, but less investigated factors related to vocational choice is the desire for security. A 40-item inventory designed specifically to measure the degree of emphasis placed upon security in vocational choice by male college upperclassmen, was developed. Results on this Security Inventory, the Edwards Personal Preference Schedule, and a Personal Preference Questionnaire to ascertain background

information were obtained from a heterogeneous group of 346 upperclassmen. It was found that emphasis upon a desire for security in choosing a job or an occupation, exists as a trait of some comprehensiveness that can be measured reliably, is positively related to the actual choice of a job situation of a secure type, and is characterized by low, but statistically significant, personality and background correlates.

#### REFERENCES

- CAPLOW, T. *The sociology of work*. Minneapolis: Univer. Minnesota Press, 1954.
- DAUGHERTY, C. R. *Labor problems in American industry*. (5th ed.) New York: Houghton-Mifflin, 1941.
- EDWARDS, A. L. *Manual of the Edwards Personal Preference Schedule*. (Rev. ed.) New York: Psychological Corporation, 1957.
- HAYEK, F. A. *The road to serfdom*. Chicago: Univer. Chicago Press, 1944.
- HOPPOCK, R. *Job satisfaction*. New York: Harper, 1935.
- LINDAHL, L. G. What makes a good job? *Personnel*, 1949, **25**, 263-266.
- LORGE, I., & BLAU, R. D. Do the better qualified compete for public jobs? *Publ. personnel Quart.*, 1942, **3**, 29-36.
- MCGREGOR, D. Conditions of effective leadership in the industrial organization. *J. consult. Psychol.*, 1944, **8**, 55-63.
- NOLAN, C. Y. Attitude differences among disparate Air Force specialties. *USAF Personnel Train. Cent. res. Rep.*, 1956, No. TN-56-88.
- RIESMAN, D., DENNY, R., & GLAZER, N. *The lonely crowd: A study of the changing American character*. (Abr. ed.) Garden City, N. Y.: Doubleday, 1953.
- ROE, ANNE. *The psychology of occupations*. New York: Wiley, 1956.
- SHARTLE, C. L. *Occupational information: Its development and application*. (2nd ed.) New York: Prentice-Hall, 1952.
- THORNDIKE, R. L., & HAGEN, E. P. Attitudes, educational programs, and job experiences of airmen who did not reenlist. *USAF Personnel Train. Cent. res. Rep.*, 1957, No. TR-57-2.
- VITELES, M. S. *Motivation and morale in industry*. New York: Norton, 1953.
- WHYTE, W. H. *The organization man*. New York: Simon & Schuster, 1956.

(Received June 21, 1961)

# THE EFFECTS OF MANNER OF SPEECH ON APPRECIATION OF SPOKEN LITERATURE

GEORGE M. GLASGOW<sup>1</sup>

*Fordham University*

Esthetic appreciation is recognized as an important objective of education and condition of artistic development. In the study and teaching of literature, there is a need for identification and measurement of elements of communication which promote the appreciation of literature. Hence, the present investigation has been designed to measure the effects of certain speech characteristics on audience appreciation of descriptive prose and lyric poetry. The problem is to determine whether or not variations in the manner of speaking the literature increase or decrease audience appreciation.

Theoretically, the problem has assumed psychological relevance with the appearance of a thoughtful survey of the literature on esthetics by Pratt (1961). He refers to the concept of "tertiary qualities" (p. 80): "the expression of art as a property closely bound with the perceptual structure" of the person who responds as an aspect of esthetic theory to be explored further. In this inquiry, the expressive qualities of elements of speech have been varied in terms of postulated "good" or desirable and "poor" or undesirable modes of presentation. The operations for exploring the esthetic experiences of human observers are the elicitation of discriminatory reactions upon hearing the varied presentations.

<sup>1</sup> The author wishes to express gratitude for the review of this manuscript by Dorothea McCarthy.

## PROBLEM

The expressive elements of speech tested were voice quality, pitch, rate, and enunciation. Within each of these elements the effects of a good manner were contrasted with those of a poor manner in terms of their influence on audience appreciation. Each speech element was treated as the only variable in one experimental unit. There were thus four experimental units. In the first unit, the effects of a good voice quality were compared with the effects of a nasal twang quality. In the second unit, varying pitch was compared with a monopitch. The third unit treated moderate rate versus fast rate; and the fourth unit, distinct enunciation versus indistinct enunciation.

A difference exists between the direct color stimuli of a painting, or the audible tones of music, and imagined color or audible stimuli which result from hearing spoken-word image symbols. There is but one hue or shade of red at any point of a picture. But the theoretical number of specific hues or shades of red which may be imagined in response to the spoken word "red" may be many and may be limited only by the spectra of one's imagination and the cumulative layers of experience. The present investigation explores the relationship between the sound spectra of voice and speech and associated variations in audiences' visual and auditory images, moods, ideas, and literary values.

Appreciation was considered as a

mental and emotional response to an object, situation, or condition in which percepts, concepts, images, and evaluations are personalized in the meanings and feelings of the appreciator. Appreciation has been viewed as self-expression. "The lyric poem is appreciated by us as an expression of our inner life" (Parker, 1920, p. 40). As self-expression, appreciation would seem to overlap "creativity," explained by Rhodes (1961, p. 305) as a "phenomenon by which a person communicates a new concept (which is the product)."

The process of appreciation may be treated analytically and synthetically. Parker (1920, pp. 188-193) has referred to the elements of literary appreciation, e.g., images, sound sensations, meanings, feelings, the rhythm of ideas; but these find embodiment and unity in the meaning of the whole esthetic experience.

The measurement of appreciation and esthetic experience has been studied by numerous investigators. Speer (1929) conducted research on the "Measurement of Appreciation in Poetry, Prose, and Art and Studies in Appreciation." Langfeld (1920) and Birkhoff (1933) evaluated the "aesthetic attitude" and esthetic measures. Recently Thurstone (1959) described a method of measurement which seems to be basically comparable to one of the methods used in the present study, as follows:

Some time ago I obtained some data to ascertain whether we can make an experimental determination of the zero point in a scale of preferences. The zero point would be that point at which the subject is neutral as to whether he wants the object or not. Any object in the scale below the neutral point is then an object which the subject dislikes.... We have made an initial analysis ... and the results are very satisfactory (p. 181).

In this investigation, good quality was a pleasant, male baritone quality. The nasal twang was produced by a constriction of the nasopharynx. Good pitch was an appropriately varying intonation. The monopitch was a sustained pitch of a piano "middle C." The good rate was 130 words per minute; the fast rate was 250 words per minute. Good enunciation was distinct but not "artificial" or overprecise; poor enunciation was moderately indistinct and portions of it were slightly slurred.

#### METHOD

The investigation consisted of two one-group experiments conducted with two similar but different groups. The second experiment was essentially a replication of the first, with the exception of changes in the order of presenting the good-poor sequences and their specimens. The purpose of repeating the experiment with a second group was to secure additional evidence and to provide opportunities for changes in the order of rotation.

In each experiment, the "good" level of a speech element, e.g., quality, was applied to one member of a pair of literary specimens and presented to an audience and their appreciation of the spoken specimen was measured. Then the "poor" level of the same speech element was applied to the paired mate of the first specimen and presented to the same audience and their appreciation of the second specimen was likewise measured. If the good level of a variable was presented first, the poor level immediately followed, and if the poor level was presented first, the good level immediately followed.

When the good level of any variable was presented in a "set," all other variables and all other recognizable speech characteristics in the same good presentation were kept good as far as was possible. When the poor level of the same variable was presented, all speech factors remained good as far as possible but the single, poor speech level. Thus, any differences between the audience's appreciation of the first and the second specimens spoken in the good and the poor speech levels were attributed to the different effects of these levels on appreciation.

*Literary Specimens.* It was recognized

that the two literary specimens used to express the two different levels of any variable should be alike in order that the same potentials of appreciation would be available on which these good and poor levels could exert their possibly varying expressive influences. Exact duplicates, however, would have entailed the use of the same literary specimen for any good-poor set, since every literary specimen is unique, and this would have involved the repetition of "content" for every second presentation in a good-poor set. Consequently, the use of a pair of similar matched specimens was considered to be preferable, conditioned with certain safeguards to enhance the value of the results.

First, two pairs of prose and poetry specimens were used for each speech element instead of the minimum of one pair. Second, a rotation was employed for the order of presentation of the good and poor levels of each element, for the speech elements themselves, and for the particular specimen in a pair used to express the good or the poor level of a speech element. It may be noted that "The rotation method . . . is frequently used when . . . there is doubt concerning the equivalence of the groups" (Good, Barr, & Scates, 1938, p. 495). In this experiment, there was doubt concerning the equivalence of the two specimens of each literary pair. Hence the three factors mentioned were rotated. Because of the nature of the experiment, it was not necessary to equate the two groups, subject against subject. Third, although the experiment with the first group was complete, the basic experiment was repeated, as indicated, in order to secure additional evidence and to provide opportunity for further rotation of the speech levels and specimens.

*Pairing of Specimens.* The matching of the literary specimens in each pair<sup>2</sup> was done by 10 experienced high school and college English teachers on the following bases: suitability of content for the grade level, similarity of literary style, potentials for appreciation, length, literary value, and relative unfamiliarity to the subjects. The pairs were designated: A-B, C-D for quality; E-F, G-H for pitch; I-J, K-L for rate; and M-N, O-P for enunciation.

*Subjects.* The two groups of subjects consisted of 90 and 136 fourth term high school sophomore girls. The groups approximated

each other to the following extent: All members of both groups were of the same grade level and sex. Only those with normal hearing, as determined by two audiometer tests, were included in either group. The mean IQ for Group I was 105.5 and for Group II, 105.1, with standard deviations of 11.9 and 10.3, respectively. There were no significant differences in the distributions of the IQ scores. English grades were also considered as possible indexes of the students' potential to appreciate English literature. Averages of four English grades yielded a mean of 78% with a range of 67-91% for Group I and 76% with a range of 63-91% for Group II. In view of the mean IQ and mean English marks, it was believed the two groups were similar samples of female high school students for purposes of this experiment.

*Judging of Specimens.* Three experienced university speech teachers agreed unanimously as to the goodness and poorness of the opposite levels of each variable. They also agreed as to the basic goodness and sameness of the other recognizable speech aspects in both presentations of each "set" for each variable. The good characteristics were judged to be esthetically pleasing, to express suitable emotional states, to sustain attention, and to provide appropriate intonation, emphasis, etc. The good manners were believed to be better than the average comparable characteristics of college students; the poor manners moderately worse.

*Recordings.* High-fidelity recordings were used in order to keep the presentations constant, to insure that the speech examples which were approved were the ones used during the procedure, and to isolate speech as the experimental factor apart from gestures and facial expression. All the speech examples were recorded by the writer.

*Rotational Sequence.* The pairs of literary specimens were presented in good-poor sets. A set was a pair of specimens with one member spoken in the good level and the other in the poor level of the same speech element. The sets were presented to Group I in two series as follows: quality, A good-B poor; pitch, G poor-H good; rate, I good-J poor; enunciation, O poor-P good. In the second series the second set of specimens for each variable was presented in the opposite set sequence, namely, quality, C poor-D good; pitch, E good-H poor; rate, K poor-L good; enunciation, M good-N poor. The sequence was the same for prose.

In the procedure with Group II, the set sequence was the reverse of that used for Group I. Specimen A was presented with poor quality and B with good quality; G with good pitch modulation and H with

<sup>2</sup> Typical of a "pair" of poetry specimens were "Love Song from New England" by Winifred Wells and "I Shall Not be Afraid" by Aline Kilmer. Typical of the prose specimens were selections from Anne Lindbergh's *North to the Orient*. Prose specimens were limited to 260 words each.

monopitch. The remaining variables followed a similar order.

*Measurement of Appreciation.* The experiment required an appreciation test which would show changes of expressed appreciation in accordance with differences in the good and poor presentations of the variables. Since such tests were not available, it was necessary to construct original tests for this study.

The tests were basically check sheets having steps representing different degrees of response. Two techniques of measuring appreciation were included in each test given. The first technique secured as indexes of appreciation the listeners' estimates of one of three degrees of the *presence* of each of five specified appreciation elements in each selection heard *plus* the extent to which the listeners *liked* each element. The three degrees of presence from which one was to be checked were strong, medium, and weak. The three steps of liking for the elements were strong, indifferent (neutral), and disliking.

The appreciation elements listed for prose were recognition of topic, clarity of thought, unity, vitality, and sense imagery. Those for poetry were recognition of theme, mood, force of emotion, sense imagery, and rhythm.

The second technique of testing secured the listeners' estimates of the *literary merit* of each whole specimen heard *plus* their personal preference for each whole specimen. Five levels of literary merit were mentioned: very superior (a masterpiece), superior, average textbook material, standards of a less important magazine, and below newspaper standards. In treating the data later the first two levels were combined into one termed "high" and the last two into one called "low." Hence, literary merit as treated became high, medium, and low. The levels of preference were also high, medium, and low.

## RESULTS

### Appreciation Elements

The first step in treating the data was the combination of each score on the "presence" of each appreciation element with the parallel score on liking in order to form a joint index of "perception" plus emotional response for each specimen. This provided five indexes for every poetry and prose selection heard by each subject.

Since each presence score could be combined with any of the three degrees of liking, there were nine possible combinations of presence and liking for each of the five elements in every selection. These were: strong presence, Number 1, with any of the three degrees of liking; medium presence, Number 2, with any of the three degrees of liking; and weak presence, Number 3, also with any of the three degrees of liking. In such a circumstance, the presence-liking ratio provided three cells of perfect agreement, namely 1:1, *high*; 2:2, *medium*; and 3:3, *low* appreciation.

The scores for the two good specimens in each two sets for each variable were averaged and the scores for each two poor specimens opposed were likewise averaged to form good and poor categories. The percentages of the scores in each of the nine presence-liking cells were computed for the good and the poor manners of each variable in prose and in poetry. These percentages showed that from 75-87% of all scores had been placed in the cells of perfect agreement: the 1:1, the 2:2, or the 3:3.

The data showed that the percentages of subjects expressing high appreciation of the individual elements were consistently greater when the speech manner was good than when it was poor. The number expressing low negative appreciation was consistently greater when the speech manner was poor.

The next step was the computation of an appreciation index which was an average of all five elements for each specimen, good or poor, heard by each subject.

These indexes for both groups of subjects were so similar that the scores of both groups in the same categories were combined from here on. The distributions of average response scores of Group I and Group II were found

TABLE 1

PERCENTAGES OF SUBJECTS IN THE HIGH, MEDIUM, AND LOW APPRECIATION CELLS BY TYPE OF SPEECH PRESENTATION OF EACH VARIABLE IN PROSE AND POETRY AND THE CHI SQUARES

Appreciation	Quality		Pitch		Rate		Enunciation	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Prose								
High positive	40	19	44	27	55	29	47	5
Mediocre	33	38	30	32	23	33	31	29
Low negative	4	17	5	15	3	15	3	53
Chi squares*	66.61		37.23		69.11		337.65	
Poetry								
High positive	41	18	42	16	44	15	44	8
Mediocre	26	34	29	29	31	35	33	32
Low negative	4	17	5	27	3	25	2	36
Chi squares*	78.15		114.32		135.85		231.78	

\* All chi squares are significant at beyond the .001 level of confidence.

to conform very closely to each other in comparable categories. Eight comparisons were made, namely, four for the presence scores and four for the preference scores. The four comparisons of presence-score distribution were composed of two prose comparisons, good and poor, and of two poetry comparisons, good and poor. The same comparisons were made for preference. In none of the eight instances, could the "null hypothesis" of "no difference" be rejected at the 5% level of confidence. The comparisons were made by means of the Kolmogorov-Smirnov two-sample test (Siegel, 1956, p. 131). No differences between the cumulative percentage distributions of Groups I and II exceeded the critical value of 18%. These findings were considered sufficient for assuming that the scores for Groups I and II were drawn from the same population and could therefore be combined.

Table 1 lists the percentages of the total number of subjects who expressed high positive (1:1), mediocre (2:2), and low negative (3:3) appreciation in each good and poor category of each

variable and the chi squares for the "independence" of the good and the poor speech versus the three levels of appreciation.

In Table 1 variations in appreciation scores accompanying variations in speech manner are evident. The number of subjects who expressed high appreciation was markedly and consistently greater when the speech manners were good than when they were poor. In prose, for example, the good-poor percentage ratios of 40:19 for high appreciation of quality through to 47:5 for enunciation reveal two to nine times as much high appreciation when the speech styles were good as when they were poor. Similar ratios exist in poetry. Moreover, low negative appreciation of prose and poetry was greater when the speech manners were poor.

There is an interesting similarity between the parallel scores for prose and poetry. For instance, the good-poor ratio for high appreciation under quality is 40:19 for prose and 41:18 for poetry; enunciation, 47:5 for prose and 44:8 for poetry. Such percentages

would seem to indicate that speech manners affect appreciation somewhat similarly even though the types of literature may differ.

### Whole Literary Specimens

The percentages of the subjects' estimates of the literary merit of, and preferences for, the whole specimens are presented in Table 2. Since the high and low levels of merit and preference tell the essential story, the middle level is omitted here. Chi squares are listed for the independence of the high, medium, and low levels of literary merit and preference, respectively, versus the good and the poor styles. Chi squares were also computed for the good and the poor frequency deviations of the combined four variables from their theoretically equal

(1:1 basis) distribution on the high appreciation and the low appreciation levels. The chi square for high appreciation was 2000.52 and for low appreciation, 2862.84. As the chi squares indicate, the differences are highly reliable.

Table 2 shows the reduction of the subjects' ratings of the literary merit and their personal preferences for the specimens occurring concomitantly with the poor style of each speech variable. Nasality, monopitch, fast rate, and indistinct enunciation notably decreased the number of persons who rated the literary merit high in comparison with the number when the speech manners were good. The number of ratings of low literary merit and of low preference for specimens was consistently greater when a poor

TABLE 2

PERCENTAGES OF SUBJECTS ASSIGNING HIGH AND LOW LITERARY MERIT  
TO LITERARY SPECIMENS AND PREFERENCES FOR THE SPECIMENS IN  
EACH MANNER OF SPEECH PRESENTATION AND THE CHI SQUARES

	Quality		Pitch		Rate		Enunciation	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Prose								
Merit: High Low Chi squares	20 31 67.76	6 56	19 36 45.89	7 57	32 20 61.14	13 41	25 29 235.14	4 81
Preference: High Low Chi squares	30 10 97.40	11 35	30 11 44.58	16 27	48 7 76.29	22 23	39 7 333.00	5 65
Poetry								
Merit: High Low Chi squares	40 24 97.26	14 52	34 22 126.77	9 56	31 25 86.20	11 53	53 13 249.80	12 63
Preference: High Low Chi squares	45 7 116.41	18 32	49 4 186.52	11 29	47 3 143.33	19 30	63 1 315.77	13 45

speech level was used. This held true for all four speech variables. Poor enunciation was associated with greater changes in appreciation than any other variable.

### Biserial Correlations

As evidence concerning the significance of the deviations of the good versus the poor levels of each speech element, biserial correlations were computed to show the extent of the differentiation between the two levels as they operate through quality, pitch, rate, and enunciation of both prose and poetry. The biserial correlation coefficient has been interpreted as a measure of the degree of divergence of the effects of the good and the poor levels from equality.

In computing these coefficients, each subject's average of the two good scores for the presence of the appreciation elements was combined with the average of the two good literary merit scores rated for each variable in order to form a composite presence score. The same was done with the scores in the poor category. Similarly, each subject's scores for preference for the appreciation elements were combined with her preferences for the whole selections in order to form a single preference score for the good and the poor categories, respectively, for each variable. These coefficients appear in

Table 3. As may be seen, the coefficients for enunciation are considerably higher than for any other speech element, a finding which suggests the greater effect of enunciation upon appreciation.

### SUMMARY

A "good" and a "poor" manner of each of four speech variables (voice quality, pitch, rate, and enunciation) were used to present 16 paired specimens of prose and 16 specimens of poetry to 226 high school sophomores. The speech manners were good quality versus nasal twang, varied pitch versus monopitch, moderate rate versus fast rate, and distinct versus indistinct enunciation. A markedly greater number of subjects expressed high appreciation of and strong preference for the selections, and also rated the selections as of high merit, when the speech manners were good than when they were poor.

The number of subjects who rated the literary merit high was greater under the good speech manner than under the comparable poor manner of every variable. The number who rated the literary merit low was greater under every poor speech manner.

The number of subjects expressing strong preferences for the whole selections was increased by every good speech manner and decreased by every poor speech manner.

Poor enunciation decreased high appreciation and increased low appreciation in terms of the comparable good manner, more than the poor manner of any other speech element.

All the chi squares are significant at beyond the .001 level of confidence.

Thus, different manners of speaking literature have different effects on audience appreciation and audience judgments of literary merit. Consequently, it is believed that speech manner is an

TABLE 3  
THE BISERIAL CORRELATION COEFFICIENTS  
FOR THE GOOD AND THE POOR LEVELS  
FOR PRESENCE AND PREFERENCE IN  
PROSE AND POETRY

Variable	Presence		Preference	
	Prose	Poetry	Prose	Poetry
Quality	.55	.70	.55	.71
Pitch	.46	.73	.44	.78
Rate	.53	.71	.55	.76
Enunciation	.95	.94	.95	.93

important factor in the educational development of potential literary appreciations and insights.

#### REFERENCES

- BIRKHOFF, G. D. *Aesthetic measure*. Cambridge: Harvard Univer. Press, 1933.
- GOOD, C. V., BARR, A. S., & SCATES, D. E. *The methodology of educational research*. New York: Appleton-Century, 1938.
- LANGFELD, H. *Aesthetic attitude*. New York: Harcourt, Brace, 1920.
- PARKER, D. H. *The principles of aesthetics*. New York: Silver Burdett, 1920.
- PRATT, C. C. *Aesthetics. Annu. Rev. Psychol.*, 1961, **12**, 71-92.
- RHODES, M. An analysis of creativity. *Phi Delta Kappan*, 1961, **42**(7), 305-310.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SPEER, R. Measurement of appreciation in poetry, prose and art and studies in appreciation. *Teachers Coll. Contrib. Educ.*, 1929, No. 362.
- THURSTONE, L. L. *The measurement of values*. Chicago: Univer. Chicago Press, 1959.

(Received June 21, 1961)

## MEASURING THE VALUE SYSTEMS OF EDUCATION PROFESSORS<sup>1</sup>

DONALD J. VELDMAN, ROBERT F. PECK, AND CARSON McGUIRE

*University of Texas*

The basic idea of this study can be applied to any situation where two or more judges make quantified assessments of people on a series of characteristics. Most factorial studies analyze the relationships between scales (*R* technique) or between subjects (*P* technique) after averaging or combining the scores given by different judges. The point of the present study, however is this: If each judge's ratings on the total population are separately factor analyzed, it becomes possible to identify his particular way of construing the characteristics under study. Furthermore, his assumptions (often unconscious) about the way he thinks the characteristics relate to one another are directly shown by the factors that emerge from his ratings. The present method resembles Kelly's (1955) "personal construct" approach, but is applicable to *any* set of judgments, regardless of their specific content.

If, upon comparison, all or most of the judges turn out the same factor structure, they may be said to show high reliability. Moreover, such agreement can be regarded as evidence of construct validity for the dimensions of human nature on which they all agree. It must be remembered, however, that agreement does not necessarily prove the common set of conceptions to be valid. All it reflects is a common frame of reference among the judges. Validation against external ev-

idence does continue to have a persuasive desirability of its own.

A very interesting possibility arises, however, when disagreement is uncovered. When comparison of the several factor-systems reveals significant differences between the individual judges, it is possible to interpret each judge's "unique" factors as a revealing projection of his own special, somewhat private conceptual world. Indeed, it is possible to identify precisely how many of his concepts he holds in common with his colleagues, and precisely how much he differs, where he does differ. Moreover, the pattern of relationships the judge expects to obtain among the different aspects of human nature is clearly specified. Whether any given judge is "right" or not, is a subject for further determination; but what each judge currently believes to be true is made clear.

When the measured characteristics include many aspects of personality which presumably are related to teaching effectiveness, and the judges are professors of education, assessing prospective teachers, the findings take on considerable practical significance (Peck, 1960). At least, what professors believe to be true about human nature is apt to influence their perception and evaluation of their students. What they believe is also likely to color their teaching and their advice to school executives.

Finally, when the measures include an estimate of the "overall mental health" of each subject, it becomes possible to determine exactly what "mental health" means to each judge, in terms of all the behavioral characteristics included in the study.

<sup>1</sup>This research is a part of the Mental Health in Teacher Education Project, College of Education, University of Texas. This project is supported by Grant No. 2M-6635 from the Training Branch, National Institute of Mental Health.

## PROCEDURE

*Rater Constructs.* Four judges experienced in personality assessment (two professors, two research associates) independently rated 53 female sophomores in education. Fifty aspects of personality and behavior were rated on seven-point scales. A rating on "overall mental health" was included.

The appraisals were based on the following battery:

- (1) Biographical Information Form
- (2) Peck Sentence Completion, 2-D (90 items)
- (3) Group Thematic Apperception Test (8 pictures)
- (4) Bown School Situations Analyzer
- (5) Educational Issues Questionnaire

Following completion of the assessments, the ratings were transferred to IBM cards and all subsequent computations were carried out with an IBM 650 data processing system.

The 50 rating scales were intercorrelated for each rater separately. The four resulting intercorrelation matrices were separately subjected to centroid factor analysis and the resulting structures were rotated by means of the normalized varimax method which maintains orthogonality among the axes.

The factors extracted in each analysis are described in Table 1, which also includes the percentage of common variance accounted for by each factor, and the loading of the overall mental health rating on each factor.

In order to obtain an objective estimate of the degree of similarity among the various rater factors, the rotated factor loadings were used as data for an intercorrelation analysis of the 25 factor variables. Table 2 indicates the primary interrater factor relationships.

*Defining Mental Health.* In Table 3 the major personality correlates of mental health are listed. These constitute the common area of agreement about the meaning of "mental health." Table 4 shows the personality variables on whose relationships to mental health the four judges differed.

## RESULTS

### Rater Constructs

It is apparent from Table 1 that the judges employ construct systems which vary considerably in degree of differentiation. (The percentages of common variance are a more reliable index of meaningful differentiation

TABLE 1  
FACTORS EMERGING IN EACH ANALYSIS,  
THE COMMON VARIANCE ASSOCIATED  
WITH EACH, AND THEIR MENTAL  
HEALTH SIGNIFICANCE

Factor	Factor description	Percentage of common variance	Mental health loading
Rater A	Creative, perceptive, independent	29	.58
	Emotionally and judgmentally restrained	17	.43
	Optimistic, relaxed, peer oriented, sociable	24	.59
	Child oriented, affectionate, indecisive	10	.12
	Sexually integrated	2	.18
Rater B	Intelligent, perceptive, independent, creative	30	.67
	Judgmentally and emotionally restrained	15	.29
	Sociable, peer oriented	9	.37
	Child oriented, affectionate, cheerful	11	.40
	Self-confident, hopeful, relaxed	5	.27
	Nondogmatic, flexible	3	.13
	Conventional, stable, expedient	2	.15
Rater C	Independent, dominant, non-conforming	33	.74
	Judgmentally and emotionally restrained	13	.41
	Affectionate, cooperative, peer oriented	19	.34
	Mild emotional reactivity	3	.02
	Imaginative, insightful	2	-.05
	Goal oriented, emotionally constricted	3	.01
	Hopeful, sociable, rigid, "adult"	2	.04
Rater D	Creative, independent, perceptive	27	.63
	Stable, emotionally and judgmentally restrained	16	.35
	Sociable, cooperative, affectionate, cheerful, peer oriented	20	.54
	Persistent, proud, "adult"	5	-.02
	Relaxed, self-confident	3	.08
	Rational, sexually integrated, goal oriented	4	.22

than is the number of factors extracted in each analysis.)

Table 2 reveals the presence of three dimensions common to the construct systems of all four raters:

TABLE 2  
PRIMARY INTERRATER FACTOR  
RELATIONSHIPS

Factor	Rater A	Rater B	Rater C	Rater D
I	B <sub>1</sub> .97	A <sub>1</sub> .97	B <sub>1</sub> .94	C <sub>1</sub> .94
II	B <sub>2</sub> .92	A <sub>2</sub> .92	A <sub>2</sub> .90	A <sub>2</sub> .92
III	C <sub>3</sub> , D <sub>3</sub> .86	A <sub>3</sub> .86	D <sub>3</sub> .91	C <sub>3</sub> .91
IV	B <sub>4</sub> .69	C <sub>3</sub> .87	D <sub>2</sub> .42	C <sub>7</sub> .55
V	C <sub>1</sub> .47	A <sub>3</sub> .78	B <sub>1</sub> .43	B <sub>5</sub> .70
VI	—	C <sub>3</sub> .66	D <sub>5</sub> .33	C <sub>1</sub> .57
VII	—	D <sub>4</sub> .37	D <sub>4</sub> .55	—

Note.—This table is a condensation of a square matrix of factor loading intercorrelations, and includes only the highest coefficients in each column of the matrix. Decimal points have been omitted.

*Common Dimension I: Creatively Intelligent Autonomy.* Factor I accounted for the most variance in all analyses. Comparisons of the descriptions in Table 1 reveal only minor differences in emphasis among the four raters, except for the absence of intellectual attributes in Rater C's cluster.

*Common Dimension II: Emotional-Judgmental Restraint.* The key traits here are virtually identical for all raters, although Rater D placed somewhat more weight on stability of mood.

*Common Dimension III: Peer Ori-*

*ented Sociability.* All four raters agree on a core cluster of traits, with wide variations in the other variables they also include in this dimension. Rater B appears to define the dimension most narrowly, while Rater D allows a large number of traits nearly equal weight.

*Unique Dimensions.* Beginning with the fourth factor in each analysis, the dimensions become more idiosyncratic. In some cases it is possible to identify a separate dimension for one rater as part of a broader dimension in another rater's system. Considering the coefficients in Table 2, on the other hand, it would be difficult to regard factors A<sub>5</sub>, B<sub>7</sub>, C<sub>4</sub>, C<sub>5</sub>, C<sub>6</sub>, D<sub>4</sub>, or D<sub>6</sub> as anything but idiosyncratic.

Factors A<sub>4</sub> and B<sub>4</sub> are quite similar and seem to represent as a separate cluster the characteristics of affection and positive orientation toward children which Raters C and D include in their third common factor. Factors B<sub>5</sub> and D<sub>5</sub> appear to be a good deal alike. They separate off one facet of Rater A's third factor—relaxation—which only he includes in Factor III. Factor B<sub>6</sub>, although somewhat unique to this rater, represents "flexibility," a major

TABLE 3  
PRIMARY CORRELATES OF MENTAL HEALTH, AVERAGED OVER FOUR RATERS

Average correlation with mental health scale	Bipolar scale description
82	Rational-altruistic character
82	Objective, accurate perception
81	Rational, objective judgment
78	Orderly, rational thought
77	Basic trust
76	Full reality awareness
75	Self-insightful
74	Appropriate goals
73	High self-esteem and self-confidence
73	Appropriate emotional expression
72	Relaxed, comfortable
71	Positive regard for family

TABLE 4  
SCALES OF UNCERTAIN SIGNIFICANCE FOR MENTAL HEALTH

Correlation with mental health scale*				Bipolar scale description	
A	B	C	D		
28	56	48	10	Rich, complex fantasy life	vs. Dull, unimaginative
54	77	54	18	Decisive, definite judgment	vs. Ambivalent, tentative judgment
38	58	09	41	Affectionate, kind	vs. Undemonstrative of affection
74	53	67	82	Genuine interest in others	vs. Self-centered
53	60	62	17	Dominant, ascendant, competitive	vs. Submissive, following, yielding
83	81	59	73	Appropriate goals	vs. Inappropriate goals
58	74	80	47	Persistent, tenacious	vs. Quitting, gives up easily
11	54	03	-13	Strong identification with children	vs. Little identification with children

\* Highest and lowest coefficients differ significantly at the .01 level.

aspect of the third dimension for Raters C and A.

Although the idiosyncratic dimensions which are unique to single raters will not be discussed here, such dimensions are of special value in identifying conceptions which the individual does not share with anyone else. This information could be quite useful, for instance, in training assessment specialists to identify and correct conceptual sets which are mainly projections and which may distort their judgment.

#### *Mental Health Definition*

Table 3 shows the way the term "mental health" was actually used. It appears that "rational-altruistic character" (Peck & Havighurst, 1960) is equivalent to "mentally healthy" for these judges. Of paramount significance are such traits as objectivity, rationality, insightfulness, and basic trust in other people. Self-confidence and appropriateness of behavior also are important components.

Table 4, on the other hand, reveals certain discrepancies among the raters' viewpoints. Raters A and D place less importance on richness of fantasy and

on persistence than do Raters B and C. Rater B de-emphasizes interest in others, but stresses identification with children, as compared to the other raters. Rater C places less importance on expression of affection and on appropriateness of goals. Rater D does not place importance on decisiveness of judgment, or dominance, to the extent that the others do. It is nonetheless interesting to note that only one minor reversal of direction occurred in the correlation coefficients.

#### *DISCUSSION*

Even where judges largely agree, as these did (the reliability of their averaged ratings was .85), the construct system of each judge still shows certain unique aspects. The method of analyzing ratings which was used here reveals each judge's system of beliefs about the structure of human nature.

If, in some other study, judges were to disagree notably, such a method of analysis would identify even more sharply each judge's unique set of concepts. As a study in the psychological beliefs of people like professors, this might be interesting in itself. Beyond

this, however, clear, precise identification of the specific points at which differences exist could conceivably lead to clearer, more conscious, and more rational discussion among judges, until consensus was reached or they agree to disagree. In either case, this kind of analysis should prove a healthy reminder that the arguments are not necessarily about the facts of human nature, but may sometimes arise from differing individual construct systems. The construct systems, in this case, are laid out on the table, for all to see.

Such a procedure, needless to say, is not limited to studies of personality. It could as usefully be applied to studies of ethical values, beliefs about educational practices—indeed, any issue important enough to warrant some planned scaling of judgments, and a moderate amount of computer time.

#### SUMMARY

A generalized method was developed for identifying the individual construct systems of judges of human behavior. The method is that of separately factor analyzing each judge's ratings on whatever scales are used in an assessment of people. Thereafter, correlation of the factors generated by several judges indicates the nature and degree of similarities and differences in the way the different judges define "human nature," and the way in which they believe different aspects of behavior are related to one another.

In the present instance, four educational psychologists rated 53 sopho-

more coeds, on 50 personality variables and on a scale of "overall mental health." The ratings were based on qualitative analysis of a battery including a biographical form, the Peck Sentence Completion (2-D), group TAT, Bown School Situations Analyzer, and an Educational Issues Questionnaire. The average interjudge correlation for all ratings was .85. Even so, some significant differences were found in the construct systems of the four judges. Each showed certain unique concepts about the structure of personality. Each differed at certain points, about the personality correlates of mental health, although there was agreement among all the judges that 12 personality variables are closely related to mental health.

The general logic of this procedure can be applied to *any* kind of judgments, regardless of the specific variables involved. It thus offers a generally applicable way of measuring individual construct systems, it appears, in *any* field where human beings make judgments. To differentiate it from *R*, *P*, and *Q* techniques, perhaps it might be called "*J* technique."

#### REFERENCES

- KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955. 2 vols.
- PECK, R. F. Personality patterns of prospective teachers. *J. exp. Educ.* 1960, **29**, 169-175.
- PECK, R. F., & HAVIGHURST, R. J. *The psychology of character development*. New York: Wiley, 1960.

(Received June 24, 1961)



# **GRADUATE EDUCATION IN PSYCHOLOGY**

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:

Anne Roe, *Chairman,*

and

John W. Gustad, Bruce V. Moore,  
Sherman Ross, and Marie Skodak

Price \$1.50

**THE AMERICAN  
PSYCHOLOGICAL ASSOCIATION**  
Dept. Grad  
1333 Sixteenth Street, N.W.  
Washington 6, D.C.

# *Psychology and Mental Health*

A report of the Institute on Education and Training  
for Psychological Contributions to Mental Health,  
held at Stanford University in August, 1955.

*Edited by CHARLES R. STROTHER*

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: \$1.75

---

*Order from:*

American Psychological Association, Inc.  
1333 Sixteenth Street, N. W.  
Washington 6, D. C.